Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity

Isabel Nocedal[1], Michael T. Laub[1,2,3]

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]correspondence: laub@mit.edu, 617-324-0418

Abstract

Gene duplication is crucial to generating novel signaling pathways during evolution. However, it remains unclear how the redundant proteins produced by gene duplication ultimately acquire new interaction specificities to establish insulated paralogous signaling pathways. Here, we used ancestral sequence reconstruction to resurrect and characterize a bacterial two-component signaling system that duplicated in α-proteobacteria. We determined the interaction specificities of the signaling proteins that existed before and immediately after this duplication event and then identified key mutations responsible for establishing specificity in the two systems. Just three mutations, in only two of the four interacting proteins, were sufficient to establish specificity of the extant systems. Some of these mutations weakened interactions between paralogous systems to limit crosstalk. However, others strengthened interactions within a system, indicating that the ancestral interaction, although functional, had the potential to be strengthened. Our work suggests that protein-protein interactions with such latent potential may be highly amenable to duplication and divergence.

Introduction

Protein-protein interactions are critical for most cellular functions, including signal transduction pathways. Notably, many protein interaction domains in signaling proteins are members of large paralogous families. For example, in mammals there are dozens of SH2, SH3, and PDZ domains that each mediate a variety of protein-protein interactions (Pawson, 2004) These paralogous proteins arise through a process of gene or domain duplication, which enables organisms, on an evolutionary timescale, to rapidly expand their signaling repertoires (Alm et al., 2006; Corrochano et al., 2016). However, the use of paralogous proteins and domains comes at a cost, requiring cells to avoid deleterious crosstalk between highly similar proteins and domains (Bradley and Beltrao, 2019; Capra et al., 2012; Siryaporn and Goulian, 2008; Zarrinpar et al., 2003). How the process of duplication and divergence unfolds at a molecular level to ensure the specificity of paralogous signaling proteins is not clear.

The precise mutations required to produce highly specific protein-protein interactions upon duplication is largely unexplored, in part because the underlying duplication events that produced most extant paralogs were ancient events. One approach to tackling this problem involves ancestral protein reconstruction, which uses the phylogenies of extant proteins to infer the sequences of ancient proteins (Hochberg and Thornton, 2017). This approach has been powerfully applied to examine the evolution of protein-ligand interactions (Voordeckers et al., 2012), such as steroid hormones and their receptors (Bridgham et al., 2009, 2006), transcription factor-DNA interactions (Baker et al., 2013; McKeown et al., 2014; Starr et al., 2017), and protein-drug interactions (Wilson et al., 2015). There have been fewer studies applying ancestral protein reconstruction to protein-protein interactions (Holinski et al., 2017; Laursen et al., 2020; Wheeler et al., 2018; Wheeler and Harms, 2021), with most focusing on resurrecting the mutations that impact protein oligomerization (Hochberg et al., 2020; Pillai et al., 2020).

We sought to understand how paralogous protein-protein interactions arise through duplication and divergence, focusing on bacterial two-component signaling pathways. Two-component signaling systems typically consist of a sensor histidine kinase (HK) that autophosphorylates upon signal recognition, and then transfers a phosphoryl group to a cognate response regulator (RR) that can trigger an intracellular response, frequently through changes in gene expression (Figure 1A) (Buschiazzo and Trajtenberg, 2019; Capra and Laub, 2012). Most bacteria encode dozens of these systems, with each system usually insulated from every other paralogous system (Galperin, 2005; Koretke et al., 2000; Skerker et al., 2005). Crosstalk between systems appears relatively rare, and has been shown to produce fitness defects when introduced

artificially (Capra et al., 2012), likely due to the decreased ability of a given HK to activate a given RR (Deeds, 2014). The specificity of the HK-RR interaction is determined primarily through molecular recognition, with a relatively small number of amino acids in both the histidine kinase and response regulator promoting the cognate interaction and preventing unwanted crosstalk with non-cognate proteins (Capra et al., 2012; Skerker et al., 2008).

Despite their prevalence in bacterial genomes and the prior identification of the key specificity-determining residues, it remains unclear how crosstalk between recently duplicated HK-RR pairs is eliminated to establish two insulated pathways. While previous work demonstrated how extant proteins can be rewired to recognize different substrates (Capra et al., 2010; Skerker et al., 2008), it is unclear if duplicated systems resolve crosstalk in a similar way. In particular, it is not known how many of the four proteins involved must acquire mutations to insulate the two protein-protein interfaces. In principle, there are two general models for how interacting proteins could evolve specificity after a duplication event (note that HK-RR systems are typically co-operonic and likely duplicate as an operon) (Figure 1B). In the first model, mutations could occur in both proteins of one system to retain their compatibility with each other while rendering them incompatible with proteins in the other system, leading to insulated pathways. This model is akin to neofunctionalization, in which a new, unique interface evolves in one paralogous system. In the second model, changes in all four proteins, perhaps each of smaller magnitude, could be required to insulate these systems. This model could represent neofunctionalization of both systems or subfunctionalization, where the functions of an ancestral protein are divided between paralogs.

Here, we use ancestral protein reconstruction to resurrect extinct HK and RR proteins that existed prior to a two-component system duplication event. By characterizing these ancestors, as well as mutational intermediates that descended from them, we elucidate the likely mutational trajectories taken by these proteins that resulted in the insulation of their protein-protein interfaces. We find that just three mutations can largely account for the establishment of specificity in the two pathways. Unexpectedly, these three mutations occur in the HK of one system and the RR of the other. The mutations that arise in one of the HKs serve mainly to prevent crosstalk to the RR of the other system whereas the mutation that arises in the RR serves both to prevent crosstalk and promote interaction with its cognate HK, suggesting that the HK-RR system that existed pre-duplication had the potential for faster phosphotransfer. By exploiting this latent ability to improve the interaction, along with the emergence of mutations that directly block crosstalk, these two HK-RR systems evolved specificity. Thus, our results

reveal the likely mutational trajectory responsible for the rapid establishment of specificity in paralogous proteins immediately post-duplication.

Results

EnvZ/OmpR has undergone duplication and diversification in α-proteobacteria

EnvZ-OmpR is a widespread two-component signaling system that has been best characterized in *E. coli* (Cai and Inouye, 2002). Many, though not all, α-proteobacteria contain two paralogous EnvZ-OmpR systems that appear to descend from a duplication that occurred in a basal α-proteobacterium (Figure 1C). In *Caulobacter crescentus* these two systems are CC1181-1182 and CC2932-2931, which we hereafter refer to as EnvZ1-OmpR1 and EnvZ2-OmpR2. These systems share less than 50% sequence identity, with particularly high divergence in the sensory domains (Figure 1 – Figure Supplement 1A). The duplication of EnvZ-OmpR in α-proteobacteria correlates with an absence of the related two-component signaling systems RstAB and CpxAR, which are highly conserved in the γ-proteobacteria (Figure 1C).

To determine the specificity of the EnvZ1-OmpR1 and EnvZ2-OmpR2 systems at the level of phosphotransfer, we used biochemical assays with purified proteins *in vitro*. For EnvZ1 and EnvZ2, we purified the cytoplasmic, catalytic domains (DHp and CA) of each kinase fused to an N-terminal MBP-His_6 tag. For OmpR1 and OmpR2, we purified each full-length response regulator harboring an N-terminal Trx-His_6 domain. To assess phosphotransfer, a given histidine kinase was first incubated with [γ-^{32}P]-ATP to drive autophosphorylation and then mixed at a 1:4 molar ratio with a response regulator of interest. Samples taken at various time points were examined by SDS-PAGE. At t=0, before adding a response regulator, there was a single band corresponding to the autophosphorylated kinase. At subsequent time points, a second band appeared as the kinase transferred its phosphoryl group to the response regulator; efficient transfer eventually led to depletion of the autophosphorylated kinase. Histidine kinases are typically bifunctional, such that when not autophosphorylated they can drive the dephosphorylation of a response regulator. This activity typically occurs on a slower timescale than phosphotransfer, but explains why the bands corresponding to phosphorylated response regulator often decreased at later time points in our assays (Figure 1D).

Using these *in vitro* phosphotransfer assays, we first tested the specificity of each paralogous system from *C. crescentus*. EnvZ1 transferred rapidly to OmpR1, its cognate response regulator, with phosphorylated OmpR1 detected after 10 seconds and autophosphorylated EnvZ

depleted by 5 minutes (Figure 1D, top left). Autophosphorylated EnvZ1 also transferred to OmpR2, but less efficiently as indicated by the slower accumulation of phosphorylated OmpR2 and slower depletion of autophosphorylated EnvZ1, with nearly full depletion occurring only by 30 minutes (Figure 1D, top right). Similar patterns were observed for EnvZ2, which transferred very rapidly to OmpR2, but quite slowly to OmpR1 (Figure 1D).

To compare the rates of transfer from different histidine kinases, we quantified the level of autophosphorylated kinase in each phosphotransfer reaction over time (Figure 1E). The rate at which the autophosphorylated kinase decreases is a proxy for the rate of phosphotransfer. (Each histidine kinase is stably phosphorylated over 30 minutes (Figure 1 – Figure Supplement 2)). By this measure, each EnvZ paralog exhibited a clear preference for its cognate response regulator (Figure 1E). Although EnvZ1 shows a weaker preference for its cognate response regulator in these assays than does EnvZ2, EnvZ1 showed a strong preference for OmpR1 in an *in vitro* competition assay (Figure 1 – Figure Supplement 1B), and even modest substrate preferences *in vitro* can result in significant *in vivo* insulation (Capra et al., 2012; McClune et al., 2019). We conclude that each histidine kinase has a preference for its cognate, co-operonic response regulator compared to the paralogous response regulator. Further, these results indicated that since the duplication event that created the EnvZ-OmpR paralogs in α-proteobacteria, each protein-protein interaction has diverged to generate paralog specificity.

Ancestral protein reconstruction reveals early acquisition of paralog specificity

To determine the evolutionary trajectory that resulted in the diversification and phosphotransfer insulation of the EnvZ-OmpR paralogs in α-proteobacteria, we used ancestral protein reconstruction to infer the sequences of the ancestral proteins. A maximum likelihood phylogeny was inferred for 200 matched pairs of cognate histidine kinase-response regulators from the EnvZ-OmpR family and other closely related two-component signaling system families (Figure 2A; full phylogeny in Figure 2 – Figure Supplement 1). This matched HK-RR phylogeny was found to be largely concordant with phylogenies based on HK or RR sequences alone (Figure 2 – Figure Supplement 2). Based on the matched HK-RR phylogeny, we identified the maximum *a posteriori* EnvZ (catalytic domains only) and OmpR sequences immediately prior to the duplication, and immediately after the duplication (Figure 2B-C; full alignments in Figure 2 – Figure Supplement 3A-B). This duplication event was quite ancient, having occurred near the origin of the α-proteobacteria ~1900 million years (Ma) ago (Wang and Luo, 2021), and these ancestral sequences share only ~50% identity with the extant proteins in *C. crescentus* (Figure 3 – Figure Supplement 2A-B). The last common ancestral proteins from which both EnvZ-OmpR

paralogs descend will be referred to as ancHK and ancRR, while ancHK1-ancRR1 refers to the ancestor of all EnvZ1-OmpR1 proteins and ancHK2-ancRR2 is the ancestor of all EnvZ2-OmpR2 (Figure 2A). Each ancestor was reconstructed with high confidence, with mean posterior probabilities > 0.8 (ancHK = 0.85, ancRR = 0.88, ancHK1 = 0.87, ancRR1 = 0.91, ancHK2 = 0.81, ancRR2 = 0.85; Figure 3 – Figure Supplement 2C). Furthermore, the sequences generated from the matched phylogeny were highly similar to those generated from an ancestral reconstruction of HK or RR sequences individually (Figure 2 – Figures Supplements 4 and 5). Each of the reconstructed ancestral sequences was cloned, expressed, and then purified, as above.

We first tested our ancestral proteins for activity in our *in vitro* phosphotransfer experiments, and found that all show clear activity in this assay, indicating that the inferred ancestors represent functional histidine kinases and response regulators. Importantly, we observed transfer from ancHK to ancRR (Figure 2D), indicating that even for our most ancient reconstructions we generated proteins capable of interacting and engaging in a productive phosphotransfer event.

We next sought to determine whether specificity, at the level of phosphotransfer, had emerged immediately after the duplication event in ancHK1-ancRR1 and ancHK2-ancRR2, systems that share ~70% identity (Figure 3 – Figure Supplement 2B). To do so, we measured phosphotransfer *in vitro* from autophosphorylated ancHK1 and ancHK2 to ancRR1 and ancRR2 (Figure 3A and 3E). We found that each ancestral kinase robustly phosphorylated its reconstructed cognate partner, with complete transfer for ancHK1-ancRR1 after 5 minutes and almost complete transfer for ancHK2-ancRR2 after 5 minutes. In contrast, each kinase showed slower transfer to the non-cognate regulator. Similar to their orthologous counterparts in *C. crescentus*, ancHK2 showed a stronger cognate preference, with very little transfer to ancRR1 observed at 5 minutes. While ancHK1 transfers more rapidly to its non-cognate regulator than ancHK2 transfers to its non-cognate regulator, a clear preference was still observed after 5 minutes, with autophosphorylated ancHK1 fully depleted after mixing with ancRR1 but not fully depleted when mixed with ancRR2. These results indicated that phosphotransfer specificity was established in ancHK-ancRR1 and ancHK-ancRR2 shortly after the duplication of ancHK-ancRR.

To determine which of the proteins acquired mutations that prevented crosstalk between paralogous systems, we first examined the phosphotransfer properties of ancHK and ancRR. We found that ancHK transferred robustly to ancRR (Figure 2D), as well as to ancRR1 and ancRR2 (Figure 3B), and that ancHK1 and ancHK2 both transferred robustly to ancRR (Figure

3D). We then compared the ability of ancestral histidine kinases pre- and post-duplication to transfer to a given post-duplication response regulator. We found that transfer from either ancHK or ancHK1 to ancRR2 was similar, (Figure 3C, right) suggesting that ancHK1 did not acquire mutations to prevent crosstalk with ancRR2. In contrast, we found that ancHK2 transferred to ancRR1 much more slowly than did ancHK (Figure 3C, left), suggesting that ancHK2 must have acquired mutations post-duplication that prevent crosstalk with ancRR1 (all transfers shown together in Figure 3 – Figure Supplement 1).

We then compared the ability of ancRR, ancRR1, and ancRR2 to be phosphorylated by ancHK1, finding that ancRR was not phosphorylated more rapidly than ancRR2 (Figure 3E, left), indicating that ancRR2 did not acquire mutations that prevent crosstalk with ancHK1. We also compared the ability of ancRR, ancRR1, and ancRR2 to be phosphorylated by ancHK2 (Figure 3E, right). In this case, we find that ancRR1 was phosphorylated much more slowly than ancRR, indicating that ancRR1 must have acquired mutations that prevent crosstalk with ancHK2. These findings were robust to phylogenetic uncertainty, as we observed a similar pattern with the alternative ancestors (Figure 3 – Figure Supplement 2D-E). We concluded that of the two response regulators produced by a duplication of ancRR only ancRR1 acquired mutations promoting insulation of the two paralogous pathways.

To quantify the change in specificity among ancestral proteins, we measured initial rates of phosphotransfer to estimate ratios of specificity constants. For a given histidine kinase, the ratio of specificity constants (k_{cat}/k_M) for two response regulators represents an approximate measure of substrate specificity, and likewise for a given response regulator the ratio of transfer from two different histidine kinases represents an approximate measure of specificity. Comparing the ratios of specificity constants of the ancestral histidine kinases (Figure 3F), we found that both ancHK and ancHK1 showed little kinetic preference, whereas ancHK2 showed an ~28-fold kinetic preference for ancRR2 relative to ancRR1, supporting the idea that the kinetic preference of ancHK1 did not change significantly post-duplication while that of ancHK2 did. For the response regulators, we observed the opposite pattern, with both ancRR and ancRR2 showing modest kinetic preferences for transfer from ancHK2 (~1.5 and 1.8-fold respectively) and ancRR1 showing a strong preference for ancHK1 (~4.5-fold) (Figure 3F). Taken all together, our results indicate that just two of the four paralogs, ancHK2 and ancRR1, acquired mutations that significantly alter their protein-protein interaction specificity in order to prevent crosstalk between the paralogous systems.

Reconstructing ancestral proteins is inherently probabilistic, and there is a degree of uncertainty associated with any reconstructed protein. To ensure that our conclusions were robust to this uncertainty, we reconstructed "AltAll" alternative sequences for the six pre- and post-duplication ancestors using a previously described method (Eick et al., 2016). In short, for every position at which multiple residues had posterior probabilities > 20%, the second most likely residue was included. These alternative ancestors were then tested for their ability to transfer to each other (Figure 3 – Figure Supplement 2D). Some of the alternative ancestors transferred more slowly than the primary ancestors. However, as with the primary ancestors, we found that just two of the alternative ancestors, ancHK2-alt and ancRR1-alt, showed significantly different transfer specificity when compared to the pre-duplication ancestors (Figure 3 – Figure Supplement 2E-F). This finding supports our conclusion that mutations in just two of the four paralogs were responsible for the insulation of these pathways.

A small set of mutations was sufficient to insulate ancestral paralogs

To identify the individual mutations responsible for the change in specificity of ancHK2, we compared the sequence of ancHK to that of ancHK1 and ancHK2, focusing on six positions previously shown to strongly covary between histidine kinases and response regulators and to dictate the specificity of *E. coli* EnvZ-OmpR(Capra et al., 2010). Only two of these positions differ between ancHK and ancHK2: positions 27 and 29, which have changed from an arginine and glutamate to a glutamine and alanine, respectively (Figure 4A). To determine if these mutations are paralog-specific, and thus likely to be important in insulating these systems, we compared the amino acids at these two positions in all identified extant EnvZ1 and EnvZ2 orthologs (a much larger set of sequences than was used for our ancestral reconstructions). This analysis indicated that both positions are indeed strongly paralog specific. At position 27, arginine is present in >90% of 1,886 EnvZ1 sequences but in none of the 822 EnvZ2 sequences, for which >90% of sequences feature either glutamine, glutamate, or serine (Figure 4B). At position 29, the negatively charged residues glutamate and aspartate are present in >90% of EnvZ1 sequences but <10% of EnvZ2 sequences, where alanine is present in >85% of sequences (Figure 4B).

A similar analysis was performed to identify possible causal mutations in the evolution of ancRR1 specificity. Only two of the key positions were found to differ between ancRR and ancRR1 (Figure 4C), and only one of these, position 11, showed broad paralog specificity (Figure 4D). At this position, arginine is present in >90% of OmpR2 sequences and <1% of OmpR1 sequences. Instead, negatively charged glutamate and aspartate are present in >60%

of OmpR1 sequences (Figure 4D). Importantly, all three of these potential key residues (27 and 29 in the kinase and 11 in the response regulator) were well supported positions in the reconstructed ancestors (Figure 2B-C), with none of them meeting the criteria for alternative reconstruction in the “AltAll” alternative ancestors.

To determine if these positions were responsible for the insulation of the paralogous systems, we tested the effect of substitutions at these positions in ancHK and ancRR. We first introduced the substitutions R27Q and E29A separately and together into ancHK and measured phosphotransfer to ancRR1 (Figure 4E). Relative to the parental protein, ancHK, both individual substitutions slowed transfer to ancRR1, with significantly less transfer observed at 2 minutes. When combined, these two substitutions decreased transfer to ancRR1 further, with a rate of transfer now comparable to that observed with ancHK2 (Figure 4E, 4H). These substitutions did not have a significant effect on transfer to ancRR2 (Figure 4 – Figure Supplement 1A-B). Thus, these two substitutions alone are sufficient to slow transfer from ancHK to ancRR1 and likely account for the major changes in ancHK2 that occurred post-duplication to help drive the insulation of the two paralogous pathways.

We then tested the effect of the substitution R11E in ancRR on transfer from ancHK2 (Figure 4F). We found that introducing this single substitution into ancRR was sufficient to significantly slow transfer from ancHK2, with a rate of transfer very similar to that seen for ancHK2 to ancRR1 (Figure 4F, 4I). Finally, we examined all three of these substitutions together by testing transfer from ancHK(R27Q, E29A) to ancRR(R11E), finding very slow transfer (Figure 4G), as seen with ancHK2 and ancRR1. Together, these results demonstrate that just three mutations – two in ancHK and one in ancRR – are sufficient to confer specificity to these EnvZ-OmpR paralogs at the level of phosphotransfer. Further, these results indicate that changes in just two of the four proteins, ancHK2 and ancRR1, which are notably not cognate partners, were required to insulate these systems (Figure 5C).

Ancestral interaction was not optimized for rapid phosphotransfer

Although our results are sufficient to explain how ancHK2 developed paralog specificity, it remained unclear how ancHK1 developed its specificity after duplication. While ancHK1 exhibits more crosstalk than ancHK2, it does exhibit a slight kinetic preference for ancRR1 over ancRR2 (Figure 3A). Because there were no differences in the six strongly co-varying residues between ancHK and ancHK1 or between ancRR and ancRR2 (Figure 4A), we hypothesized that mutations in ancRR1 must have been responsible for this specificity change. Indeed, when we introduced the substitution R11E into ancRR, we observed slower transfer from ancHK2, as

already noted (Figure 4F, 4I), as well as faster transfer from both ancHK1 and ancHK (Figure 5A-B). This result suggests that a single mutation in ancRR was sufficient to both improve the ancestral interaction and help prevent crosstalk with the new paralog ancHK2 (Figure 5C). This finding further suggests that the ancestral ancHK-ancRR interaction was not optimized for the most rapid possible phosphotransfer, and that ancHK1 evolved a preference for ancRR1 by simply improving the ancHK1-ancRR1 interaction such that this transfer outcompetes crosstalk to ancRR2.

To better understand why the R11E mutation in ancRR might improve phosphotransfer from ancHK, we used AlphaFold2 (Jumper et al., 2021) to predict the structure of the ancHK-ancRR and ancHK-ancRR1 complexes (Figure 5D). These structures suggested that substituting an arginine at position 11 in ancRR with a glutamate enables ancRR1 to form a salt bridge with R27 in ancHK and ancHK1. The emergence of this salt bridge may explain why the R11E substitution improves the interaction of ancRR1 with ancHK and ancHK1.

Although we observed, and can largely account for, paralog specificity in the ancestors that arise shortly after the duplication event, the extant proteins in *C. crescentus* exhibit more paralog specificity (Figure 1D), suggesting that subsequent mutations further insulated these paralogous protein interfaces. In particular, we observed only a weak preference of ancHK1 for ancRR1 relative to ancRR2 (Figure 3A), while *C. crescentus* EnvZ1 has a much stronger preference for its cognate partner (Figure 1D). To identify residues in *C. crescentus* EnvZ1 that may have provided additional specificity, we compared the sequences of ancHK1 and *C. crescentus* EnvZ1 and identified two mutations in the strongly covarying residues (positions 26 and 30) that differ (Figure 6A). When we looked at the identity of these residues in close relatives of *C. crescentus*, the *Caulobacteraceae*, we find that one of these positions – position 26 – is highly conserved in this clade, with a phenylalanine in the *Caulobacteraceae* compared to a leucine in the ancHK1 ancestor (Figure 6B).

When we introduced this substitution into ancHK1, we observed faster transfer to *C. crescentus* OmpR1 (Figure 6C), recapitulating the behavior observed for *C. crescentus* EnvZ1 (Figure 6D). This mutation had a smaller effect on transfer to *C. crescentus* OmpR2 (Figure 6C, 6D), although it also slowed transfer to this non-cognate partner. This finding suggests that a leucine to phenylalanine substitution in the lineage leading from ancHK1 to EnvZ1 provided further insulation of the two paralogous systems (Figure 6E). This observation also supports the model that insulation of the two paralogous systems was accomplished primarily by improving the

cognate ancHK1-ancRR1 interaction and breaking the non-cognate ancHK2-ancRR1 interaction (Figure 5C).

Discussion

Ancestral sequence reconstruction of protein-protein interactions

Ancestral sequence reconstruction has been used to interrogate the evolutionary history of a variety of protein functions, including DNA binding (Anderson et al., 2015; McKeown et al., 2014), small molecule binding (Bridgham et al., 2009, 2006), oligomerization (Hochberg et al., 2020; Pillai et al., 2020), and enzymatic activity (Boucher et al., 2014; Howard et al., 2014). But little has been done to investigate how gene duplication events impact protein-protein interactions. Here, we showed that this technique can be used to simultaneously reconstruct two interacting proteins, and we used these ancestors along with mutational intermediates to determine the evolutionary trajectory that allowed for the generation of new, insulated protein interaction interfaces.

In the case where a two-component signaling system, usually encoded as a bicistronic operon, is duplicated, two paralogous protein interaction interfaces are generated. If these two interaction interfaces are to ultimately support two separate signaling pathways, they must acquire mutations that somehow prevent detrimental crosstalk and ensure interaction specificity. For all two-component systems, and indeed for virtually all protein-protein interactions, the mutational trajectories responsible for establishing such specificity in paralogs post-duplication are unknown. Generally speaking, specificity can be generated via either (i) mutations in one of the paralogous systems such that the interacting proteins maintain their interaction while eliminating interaction with the paralogous system or (ii) mutations in proteins from both systems to generate specificity of both paralogous interfaces. In the case of the EnvZ/OmpR duplication in α-protebacteria, our results support the latter model that mutations in both paralogous systems were required to insulate these interfaces. Somewhat surprisingly, however, we found that changes in just two proteins (ancHK2 and ancRR1) were sufficient to establish specificity, rather than changes to all four proteins. This finding suggests that both protein interaction interfaces were impacted by the duplication event, and that a new two-component signaling system was not generated entirely via neofunctionalization of one system. Instead, proteins from both systems had to change to generate signaling specificity. It remains unclear if this mode of diversification represents the norm for duplicated two-component signaling systems, or duplicated interacting proteins more generally. Further characterization of

ancestral interacting proteins will be required to determine whether this mechanism of insulation is commonplace.

A small set of substitutions was sufficient to generate insulated protein interfaces

Previous work has shown that a small number of historical substitutions can have significant consequences for the specificity of proteins that bind DNA (McKeown et al., 2014), protein binding to small molecules(Bridgham et al., 2006), and protein multimerization (Finnigan et al., 2012; Pillai et al., 2020). For two-component signaling systems, substitutions in just a few key residues of the histidine kinase or response regulator can significantly alter the specificity of their interactions and generate new synthetic insulated protein-protein interactions (Capra et al., 2010; McClune et al., 2019; Skerker et al., 2008). However, it remained unclear if such small sets of mutations were sufficient to establish specificity upon a gene duplication event, and if the ancestral mutations that drove paralog insulation involve the same residues involved in rewiring the specificities of extant two-component systems. We found that, for the EnvZ-OmpR paralogs in α-proteobacteria, just three key ancestral substitutions (R27Q and E29A in ancHK2, and R11E in ancRR1) were indeed sufficient to establish specificity upon duplication, with one additional subsequent mutation (L26F in ancHK1) sufficient to establish specificity that resembles the specificity observed in the extant *C. crescentus* systems. These findings indicate that new, insulated two-component signaling pathways can readily evolve via gene duplication and the subsequent accumulation of just a few key substitutions.

The insulation of the paralogous ancHK1-ancRR1 and ancHK2-ancRR2 interfaces was primarily accomplished by weakening the interaction between ancHK2 and ancRR1, as well as strengthening the interaction between ancHK1 and ancRR1. This evolutionary path relied on the fact that the ancestral interaction, ancHK-ancRR, was not fully optimized for rapid phosphotransfer. Although this interaction was likely fully functional prior to the duplication event, the ability of the R11E substitution in ancHK to increase the rate of transfer from ancHK (and ancHK1) allowed the diversification of these paralogous interfaces by strengthening one of the two cognate interactions. We speculate that other two-component signaling systems and other protein-protein interactions that are similarly non-optimal may be particularly well-suited to duplication and divergence. It is worth noting that we cannot rule out the possibility that the lack of optimality we see in the ancestral sequences may be due to errors in the ancestral sequence reconstruction. However, we believe this is unlikely as we have identified two independent mutations (R11E in ancRR and L26F in ancHK1) that improve the ancHK1-ancRR1 interface.

How novel protein-protein interactions evolve

Both selection and neutral drift are important in the generation of evolutionary novelty. For two-component signaling systems, it remains an open question whether the changes that result in insulation of these systems generally accumulate slowly over evolutionary time through drift, or whether strong selection drives insulation of these systems. In the case of α-proteobacterial EnvZ-OmpR systems, we found that after a gene duplication event there was a rapid change in just a few key residues that dictate protein interaction specificity. Subsequent to this burst of changes, there were many mutations that accumulated in these paralogous proteins, but these mutations do not seem to have made major contributions to the insulation of these systems. This sequence of events suggests that strong selection against crosstalk occurs immediately post-duplication followed by long periods of relative stasis in the key specificity-determining residues and neutral accumulation of changes elsewhere in these proteins.

We have determined the likely evolutionary trajectory that resulted in the diversification of the two paralogous EnvZ-OmpR paralogous systems in α-proteobacteria. In this case, a small set of mutations in two non-cognate proteins was largely sufficient to establish insulation of the two pathways. Whether similar trajectories have been followed to establish other paralogous two-component signaling pathways remains an open question, but it seems unlikely that a single model will account for the mechanism of divergence of these systems in general. Our work also focused entirely on the two systems that were produced by a duplication event. However, prior work has indicated that the avoidance of crosstalk with other, existing paralogs following a duplication event can also select for changes in specificity residues (Capra et al., 2012). Further studies of how paralogs emerge, using similar ancestral reconstruction methods as used here, promises to shed more light on the general principles and mechanisms by which two-component signaling pathways, and other protein-protein interactions found throughout biology, evolve.

Methods

Ancestral protein reconstructions

EnvZ and OmpR homologs from the ProGenomes database (Mende et al., 2019) were identified using HMMER (Eddy, 2011). Cognate histidine kinase and response regulator pairs were matched based on genome proximity, with only adjacent genes matched. Clusters of HK-RR genomic sequences, where it was difficult to identify cognate pairs, were removed from the analysis. Protein sequences were then merged into a concatenated sequence for each matched pair. A subset of representative merged sequences (200 total) were aligned with MUSCLE (Edgar, 2004), the N-terminal sensory domain from EnvZ was removed, and sequences were re-aligned with MUSCLE. The best fit evolutionary model was selected using ModelTest (Darriba et al., 2020) and the Akaike Information Criterion (LG + gamma) and a maximum likelihood phylogeny was inferred using PhyML 3.3.3 (Guindon et al., 2010). Node support was evaluated using the approximate likelihood ratio test statistic (in PhyML); tree was rooted on Actinobacteria EnvZ-OmpR homologs MprBA. Ancestral sequences were then reconstructed using the codeml package in PAML 4.8 (Yang, 2007) using the maximum likelihood phylogeny (full DNA sequences for all reconstructed ancestors in Supplementary File 1). This process was repeated with the HK-only and RR-only alignments to determine individual phylogenies (Figure 2 – Figure Supplement 2) and ancestral sequences based on these phylogenies (Figure 2 – Figure Supplements 4 and 5). To account for uncertainty in the reconstructions, ambiguously reconstructed sites were identified as those at which multiple residues had posterior probabilities > 0.2 (Eick et al., 2012). For each ancestral protein, an alternative ancestor was generated by incorporating the second highest likelihood residue at all ambiguous sites (Figure 3 – Figure Supplement 2, full sequences in Supplementary File 1).

Identification of paralog-specific and species-specific residues

To identify paralog-specific residues in a larger set of EnvZ-OmpR sequences, a merged concatenated HMMER-aligned sequence was generated for all matched protein pairs identified using HMMER and genome proximity, as described above (~11,000 total sequences). A phylogenetic tree was constructed using FastTree (Price et al., 2009) and EnvZ1-OmpR1 and EnvZ2-OmpR2 paralogs were classified based on clade identity. To identify *Caulobacteraceae*-specific EnvZ1 residues in this same set of sequences, EnvZ1 paralogs were identified that were members of the *Caulobacteraceae* based on species classification from the ProGenomes database (Mende et al., 2017).

Species tree

To determine the distribution of EnvZ-OmpR paralogs, a proteobacterial species tree was generated based on a concatenated alignment of 27 ribosomal protein genes (*rpsD, rplD, rpsC, rplF, rpsK, rplA, rplI, rpsG, rplP, rplX, rpsH, rplJ, rplK, rplT, rplM, rpsI, rplB, rplV, rpsE, rplO, rpsA, rpsB, rpmE2, rpsF, rpsT, rplU, rplQ*) (Hug et al., 2016). HMMER was used to identify and align orthologs of these genes from the ProGenomes Database. The concatenated alignment was manually trimmed to remove positions represented in < 50% of sequences and positions with < 25% conservation, and a tree was generated using FastTree, and rooted on Cyanobacteria (Supplementary File 2). EnvZ-OmpR distribution was determined by identifying matched EnvZ-OmpR pairs from the protein phylogeny described above. Visual inspection of a species tree generated in (Park et al., 2018) suggests that use of this newer tree would likely not change our results.

Protein expression and purification

Expression and purification of EnvZ and OmpR and ancestral proteins was carried out as previously described (Skerker et al., 2005). Briefly, the cytoplasmic domains (DHp and CA) of EnvZ were purified fused to an N-terminal MBP-His_6 tag; full-length OmpR was purified fused to an N-terminal Trx-His_6 domain. Both proteins were expressed in BL21(DE3) cells and purified on a Ni^{2+}-NTA column.

Phosphotransfer assays

Phosphotransfer experiments were carried out as previously described (Skerker et al., 2005). Briefly, a given histidine kinase was first autophosphorylated with [γ-^{32}P]-ATP (Perkin Elmer) for 90 minutes at 30 °C to drive autophosphorylation and then mixed at a 1:4 molar ratio with a response regulator (1 μM EnvZ, 4 μM OmpR). Reactions were incubated at 30 °C and stopped at relevant timepoints by adding 4x Laemmli buffer with 8% 2-mercaptoethanol. Products were separated by SDS-PAGE (BioRad Any kD Mini-PROTEAN TGX Gel), exposed to a phosphor screen, and quantified with a Typhoon scanner (GE Healthcare) at 50 μm resolution. A representative image of two independent experiments is shown in figures. Images were quantified using ImageQuant, with rolling ball background subtraction (radius = 200), and normalized to t=0 lane for each HK-RR pair. To quantify kinetic preferences, initial rates of phosphotransfer were determined. (Skerker et al., 2008). Initial rates were determined by measuring the rate of loss of phosphorylated kinase between 0 and 30 seconds for cognate substrates, and between 0 and 5 min for all other HK-RR pairs. For *in vitro* competition

experiments, for increased visibility of RR bands, autophosphorylated *C. crescentus* EnvZ1 was mixed with *C. crescentus* OmpR1 and OmpR2 at a 2:1 molar ratio (8 μM EnvZ, 4 μM OmpR1, 4 μM OmpR2). After exposure to a phosphor screen, the gel was stained with Coomassie brilliant blue to distinguish response regulators by size.

Protein structure prediction

The predicted structure of the ancHK-ancRR and ancHK-ancRR1 complexes was generated using AlphaFold2 (Jumper et al., 2021), modeling the histidine kinase as a homodimer and the response regulator as a monomer (Supplementary File 4). Default parameters were used (MSA method: mmseqs2, pair mode: unpaired, number of models: 5, max recycles: 3).

Additional Files

Supplementary File 1: Excel spreadsheet containing relevant strains, primers, and protein sequences.

Supplementary File 2: Newick file of proteobacteria species tree (shown in Figure 1C). Species numbers from ProGenomes database (http://progenomes.embl.de/index.cgi).

Supplementary File 3: Newick file of EnvZ/OmpR merged phylogeny used for ancestral reconstructions (shown in Figure 2 – Figure Supplement 1). Protein numbers from ProGenomes database.

Source Files

Source Data 1: Zip file of original, uncropped gel images for all phosphotransfers.

Source Data 2: Figure showing all original, uncropped gel images with pink boxes indicating region shown in figure (figure number in purple below) and raw image file that corresponds (file number in pink below).

Figure 1-source data 1: Quantified phosphotransfer values for Figure 1E.

Figure 3-source data 1: Quantified phosphotransfer values for Figure 3C.

Figure 3-source data 2: Quantified phosphotransfer values for Figure 3E.

Figure 3-source data 3: Quantified phosphotransfer values for Figure 3 – Figure Supplement 2E.

Figure 4-source data 1: Quantified phosphotransfer values for Figure 4H.

Figure 4-source data 2: Quantified phosphotransfer values for Figure 4I.

Figure 4-source data 3: Quantified phosphotransfer values for Figure 4 – Figure Supplement 1E.

Figure 5-source data 1: Quantified phosphotransfer values for Figure 5B.

Figure 5-source data 2: PDB file for AlphaFold predicted ancHK-ancRR1 complex structure shown in Figure 5D.

Figure 6-source data 1: Quantified phosphotransfer values for Figure 6D.

Acknowledgements

We thank C. McClune, I. Frumkin, and D. Ghose for valuable discussions and comments on the manuscript; B. Wang for assistance with phosphotransfers; and J. Thornton for advice on ancestral reconstructions. This work was supported by an NIH grant (1F32GM126765-01) to I.N.

Figure Legends

Figure 1: EnvZ-OmpR duplication and divergence in *C. crescentus*.

(A) Two-component signal transduction. A sensor histidine kinase autophosphorylates upon activation and transfers a phosphoryl group to a cognate response regulator to activate an intracellular response.

(B) Two models for acquisition of paralog specificity after duplication of an interacting histidine kinase (HK) and response regulator (RR). Model 1: both proteins in one system acquire compensatory mutations that maintain their interaction while preventing interaction with the other system. Model 2: all four proteins acquire mutations that prevent crosstalk between systems.

(C) Phylogenetic species tree of proteobacteria inferred from 27 ribosomal protein sequences showing distribution of EnvZ-OmpR homologs and related systems. Scale bar indicates substitutions per site.

(D) *In vitro* phosphotransfer specificity of *C. crescentus* EnvZ and OmpR paralogs. Purified histidine kinase (cytoplasmic domain only) was autophosphorylated and then mixed with a given purified response regulator and incubated for the time indicated. At t=0 a single upper band corresponds to the autophosphorylated HK. At subsequent time points, a second, lower band corresponding to the RR appears as the kinase transfers its phosphoryl group leading to depletion of the autophosphorylated HK. At longer time points phosphatase activity of the HK can lead to disappearance of the phosphorylated RR.

(E) Quantification of phosphorylated HK over time in (D). Values were normalized to t=0 for each HK-RR pair. Lines represent mean (n = 2) and dots represent independent replicates.

Figure 2: Inference of ancestral α-proteobacterial EnvZ-OmpR proteins.

(A) Simplified phylogenetic tree of merged, matched EnvZ and OmpR sequences. Number of sequences in each clade indicated. Node support indicated by approximate likelihood ratio statistic (* indicates >10, ** indicates >100). Circles represent reconstructed ancestral protein pairs. Scale bar represents substitutions per site. For complete phylogeny, see Supplemental Figure 2.

(B-C) Multiple sequence alignment of EnvZ DHp domains (B) and OmpR receiver domains (C) from extant *C. crescentus* paralogs, ancHK-ancRR, ancHK1-ancRR1, ancHK2-ancRR2, and extant *E. coli* EnvZ-OmpR sequences. Residues conserved in both *C. crescentus* paralogs and all ancestors highlighted in grey; residues previously shown to strongly covary and dicatate specificity in *E. coli* EnvZ (Capra et al., 2010) highlighted in yellow. Secondary structure elements, based on AlphaFold prediction of the ancHK-ancRR complex shown below alignment. Posterior probabilities of reconstructed ancestral sequences at these positions shown for ancHK-RR (yellow), ancHK1-RR1 (blue), and ancHK2-RR2 (green) with most likely residue indicated by respective colors, and second most likely shown in grey. Dashed white line indicates posterior probability of 0.2, the threshold for identifying sites to be alternatively reconstructed (see Supplemental Figure 4).

(D) Phosphotransfer from autophosphorylated ancHK to ancRR.

Figure 3: **AncHK2 and ancRR1 acquired new specificities post-duplication.**

(A) Phosphotransfer from ancHK1 and ancHK2 to ancRR1 and ancRR2.

(B) Phosphotransfer from ancHK to ancRR1 and ancRR2.

(C) Quantification of the phosphorylated HKs indicated over time for (A) and (B) for transfer to ancRR1 (left) and ancRR2 (right). Lines represent mean (n = 2) and dots represent independent replicates.

(D) Phosphotransfer from ancHK1 and ancHK2 to ancRR.

(E) Quantification of phosphorylated HK over time from (A) and (D) for transfer from ancHK1 (left) and ancHK2 (right) to the RRs indicated.

(F) Estimate of substrate specificity for all ancestors. The ratio of specificity constants (k_{cat}/k_M) was determined for each HK or RR using the initial rate of phosphotransfer with one protein relative to another. Blue indicates a preference for ancHK1 (for RRs) or ancRR1 (for HKs), green represents a preference for ancHK2 (for RRs) or ancRR2 (for HKs), and white indicates no preference. Numbers indicate fold-preference (ratio of specificity constants).

Figure 4: **Identification of mutations responsible for ancestral paralog insulation.**

(A) Sequences of ancHK, ancHK1, and ancHK2 for regions primarily involved in molecular recognition. Dots indicate conservation compared to ancHK. Residues previously shown to be strongly coevolving and important for specificity of *E. coli* EnvZ-OmpR highlighted in yellow, and secondary structure elements predicted by AlphaFold indicated below sequence.

(B) Sequence logos for HK positions 27 and 29 in 1,886 identified EnvZ1 paralogs and 822 identified EnvZ2 paralogs, with height indicating frequency of each amino acid.

(C) Same as (A) but for ancRR, ancRR1, and ancRR2.

(D) Same as (B) but for positions 11 and 12 in OmpR.

(E) Phosphotransfer from ancHK, ancHK with the mutations indicated, and ancHK2 to ancRR1 at 0, 2 and 5 minute timepoints.

(F). Phosphotransfer from ancHK2 to ancRR, ancRR + R11E, and ancRR1.

(G). Phosphotransfer from ancHK + R27Q + E29A to ancRR + R11E.

(H-I) Quantification of normalized phosphorylated HK from (E) and (F). Lines represent mean (n = 2) and dots represent independent replicates.

Figure 5: Identification of a mutation that enhances EnvZ-OmpR interactions.

(A) Phosphotransfer from ancHK1 (left) and ancHK (right) to ancRR, ancRR + R11E, and ancRR1 at 0, 2, and 5 minute timepoints.

(B) Quantification of normalized phosphorylated HK from (A).

(C) Model for insulation of EnvZ-OmpR paralogs in α-proteobacteria. Thickness of the black arrows indicates relative strength of a given interaction. Mutations that prevent crosstalk between paralogs indicated in pink; mutations that improve cognate interaction indicated in orange.

(D) Predicted ancHK-ancRR1 complex structure from AlphaFold2. Inset: putative salt bridge between arginine 27 in ancHK and glutamate 11 in ancRR1 indicated by dashed line.

Figure 6: Tracing the mutations that produced extant, insulated EnvZ-OmpR paralogs.

(A) Sequences of DHp domain for ancHK1 and *C. crescentus* EnvZ1 shown as in Figure 4A.

(B) Sequence logos for HK positions 26 and 30 in 1,886 identified EnvZ1 paralogs and 32 EnvZ1 paralogs from *Caulobacteraceae* species, with height indicating frequency of each amino acid.

(C) Phosphotransfer from ancHK1, ancHK1 + L26F, and *C. crescentus* EnvZ1 to *C. crescentus* OmpR1 (left) and *C. crescentus* OmpR2 (right).

(D) Quantification of normalized phosphorylated HK from (C).

(E) Expanded, simplified phylogeny of α-proteobacterial EnvZ-OmpR paralogs showing origins of key historical mutations leading to *C. crescentus* and whether they affect crosstalk (pink) and/or cognate interaction (orange).

Figure Supplement Legends

Figure 1 – Figure Supplement 1: Comparison of extant EnvZ-OmpR paralogs.

(A) (Top) Comparison of *C. crescentus* EnvZ1 (CC1181) and EnvZ2 (CC2932) domain structure, with sequence identity indicated below each domain (TM = transmembrane domain, DHp = dimerization and histidine phosphotransfer domain, HAMP = histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatases domain, CA = catalytic and ATP binding domain). (Bottom) Comparison of *C. crescentus* OmpR1 (CC1182) and OmpR2 (CC2931) domain structure, with sequence identity indicated below each domain (RD = receiver domain, DBD = DNA-binding domain).

(B) Coomassie blue stain (left) and autoradiograph image of SDS-PAGE showing phosphotransfer from *C. crescentus* EnvZ1 to mixed *C. crescentus* OmpR1 and OmpR2. Arrows indicate size of each protein as determined by Coomassie stain.

Figure 1 – Figure Supplement 2: Phosphorylation of histidine kinases alone

Phosphorylation of all described histidine kinases with no response regulator added for 30 minute timecourse.

Figure 2 – Figure Supplement 1: Phylogeny of EnvZ-OmpR paralogs and related systems.

Full phylogeny of matched, merged EnvZ and OmpR sequences and related two-component systems RstBA, CpxAR, and MprBA. Node support shown as approximate likelihood ratio statistic. Sequence identifiers from ProGenomes database (Mende et al., 2017). Reconstructed ancestral nodes indicated by colored circles. Scale bar represents substitutions per site.

Figure 2 – Figure Supplement 2: HK-only and RR-only phylogenies

Simplified phylogenetic tree of EnvZ (A) and OmpR sequences (B) based on non-merged multiple sequence alignments. Number of sequences in each clade indicated. Node support indicated by approximate likelihood ratio statistic (* indicates >10, ** indicates >100). Scale bar represents substitutions per site.

Figure 2 – Figure Supplement 3: Multiple sequence alignments of EnvZ and OmpR proteins.

(A-B) Multiple sequence alignment of EnvZ (DHp and CA domains) (A) and OmpR full protein (B) from extant *C. crescentus* paralogs, ancHK/ancRR, ancHK1/ancRR1, and ancHK2/ancRR2, and extant *E. coli* EnvZ and OmpR sequences. Residues conserved in both *C. crescentus* paralogs and all ancestors highlighted in grey; residues previously shown to strongly co-vary and dicatate specificity in *E. coli* EnvZ (Capra et al., 2010) highlighted in yellow. Secondary structure elements, based on AlphaFold prediction of ancHK-ancRR complex, shown below alignment. Posterior probabilities of reconstructed ancestral sequences at these positions shown for ancHK-ancRR (yellow), ancHK1/ancRR1 (blue), and ancHK2/ancRR2 (green) with most likely residue indicated by respective colors, and second most likely shown in grey. Dashed white line indicates posterior probability of 0.2, the threshold for identifying sites to be alternatively reconstructed (see Supplemental Figure 4).

Figure 2 – Figure Supplement 4: Alignment of ancestral HK sequences reconstructed using alternative techniques

Comparison of ancHK (A), ancHK1 (B), and ancHK2 (C) sequences reconstructed using merged HK-RR alignment (ancHK, ancHK1, ancHK2), HK-only alignment (ancHK*, ancHK1*, ancHK2*), and merged HK-RR alignment with alternative “AltAll” residues (ancHKalt, ancHK1alt, ancHK2alt). Positions that differ between original ancestor (first line) and HK-only ancestor (second line) indicated with asterisk; positions that differ between original ancestor (first line) and AltAll ancestor (third line) indicated with asterisk.

Figure 2 – Figure Supplement 5: Alignment of ancestral RR sequences reconstructed using alternative techniques

Comparison of ancRR (A), ancRR1 (B), and ancRR2 (C) sequences reconstructed using merged HK-RR alignment (ancRR, ancRR1, ancRR2), RR-only alignment (ancRR*, ancRR1*, ancRR2*), and merged HK-RR alignment with alternative “AltAll” residues (ancRRalt, ancRR1alt, ancRR2alt). Positions that differ between original ancestor (first line) and RR-only

ancestor (second line) indicated with asterisk; positions that differ between original ancestor (first line) and AltAll ancestor (third line) indicated with asterisk.

Figure 3 – Figure Supplement 1: All Phosphotransfer Reactions Compared

(A) Comparison of all 5-timepoint phosphotransfer reactions from ancHK, ancHK1, ancHK2, *Cc* EnvZ1, *Cc* EnvZ2, and ancHK1 + L26F to ancRR, ancRR1, ancRR2*, Cc* OmpR1*,* and *Cc* OmpR2. Timepoints shown are 0, 10 sec, 30 sec, 5 min, 30 min.

(B) Comparison of all 3-timepoint phosphotransfer reactions from ancHK, ancHK1, ancHK2, ancHK + R27Q, ancHK + R27Q + E29A, and ancHK + E29A to ancRR, ancRR1, ancRR2, and ancRR + R11E. Timepoints shown are 0, 2 min, 5 min.

Figure 3 – Figure Supplement 2: Ancestral protein reconstruction details.

(A) Number of amino acid changes between various ancestors and extant *C. crescentus* proteins for EnvZ (left) and OmpR (right). Total reconstructed HK length = 206 amino acids, RR length = 227 amino acids.

(B) Sequence identity comparison between various ancestors and extant *C. crescentus* proteins for EnvZ (left) and OmpR (right).

(C) Distribution of posterior probabilities of maximum a posteriori states for all reconstructed ancestral proteins, with mean probability indicated.

(D) Phosphotransfer for alternatively reconstructed ancestors, incorporating the second most probable residue at all sites where the probability for this residue was > 0.2. Alternative ancestors show similar properties to primary ancestors, with the exception of slower transfer to ancRR and ancRR1, perhaps due to decreased stability of these reconstructed proteins.

(E) Quantification of normalized phosphorylated HK from (D).

(F) Estimate of substrate specificity for alternative ancestors (left) and primary ancestors (right; copied from Figure 3F for comparison here). Blue indicates a preference for ancHK1-ancRR1, green indicates a preference for ancHK2-ancRR2, and white indicates no preference. Black numbers indicate fold-preference (ratio of specificity constants), while pink numbers indicate ratios of fold-preference. Ratios of fold-preference are consistent between alternative ancestors and primary ancestors, while absolute values differ most notably between HKs, with all three alternative HKs having stronger preferences for ancRR2 due to slower overall transfer observed for ancRR-alt and ancRR1-alt.

Figure 4 – Figure Supplement 1: Phosphotransfer analysis of mutations impacting paralog specificity.

(A) Phosphotransfer from the histidine kinases indicated to the response regulators indicated at 0, 2, and 5 minute timepoints.

(B) Quantification of normalized phosphorylated ancHK, ancHK2, and mutants during transfer to ancRR2.

Figure 5 – Figure Supplement 1: ancHK-ancRR1 AlphaFold structure confidence

Per-residue local Distance Difference Test (lDDT) for ancHK-ancRR1 complex. Positions 1-211 are ancHK monomer 1, positions 211-422 are ancHK monomer 2, and positions 422-600 are ancRR1. Best 5 models shown; rank 1 model shown in Figure 5.

References

Alm E, Huang K, Arkin A. 2006. The Evolution of Two-Component Systems in Bacteria Reveals Different Strategies for Niche Adaptation. *PLoS computational biology* 2:e143-14. doi:10.1371/journal.pcbi.0020143

Anderson DW, McKeown AN, Thornton JW. 2015. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *Elife* 4:e07864. doi:10.7554/elife.07864

Baker CR, Hanson-Smith V, Johnson AD. 2013. Following Gene Duplication, Paralog Interference Constrains Transcriptional Circuit Evolution. *Science (New York, NY)* 342:104–108. doi:10.1126/science.1240810

Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* 3:e02304. doi:10.7554/elife.02304

Bradley D, Beltrao P. 2019. Evolution of protein kinase substrate recognition at the active site. *PLoS biology* 17:e3000341-25. doi:10.1371/journal.pbio.3000341

Bridgham JT, Carroll SM, Thornton JW. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science (New York, NY)* 312:97–101. doi:10.1126/science.1123348

Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515–519. doi:10.1038/nature08249

Buschiazzo A, Trajtenberg F. 2019. Two-Component Sensing and Regulation: How Do Histidine Kinases Talk with Response Regulators at the Molecular Level? *Annual review of microbiology* 73:annurev-micro-091018-054627-22. doi:10.1146/annurev-micro-091018-054627

Cai SJ, Inouye M. 2002. EnvZ-OmpR Interaction and Osmoregulation in Escherichia coli *. *J Biol Chem* 277:24155–24161. doi:10.1074/jbc.m110715200

Capra EJ, Laub MT. 2012. Evolution of Two-Component Signal Transduction Systems. *Annual review of microbiology* 66:325–347. doi:10.1146/annurev-micro-092611-150039

Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, Laub MT. 2010. Systematic Dissection and Trajectory-Scanning Mutagenesis of the Molecular Interface That Ensures Specificity of Two-Component Signaling Pathways. *PLoS genetics* 6:e1001220-14. doi:10.1371/journal.pgen.1001220

Capra EJ, Perchuk BS, Skerker JM, Laub MT. 2012. Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. *Cell* 150:222–232. doi:10.1016/j.cell.2012.05.033

Corrochano LM, Kuo A, Marcet-Houben M, Polaino S, Salamov A, Villalobos-Escobedo JM, Grimwood J, Álvarez MI, Avalos J, Bauer D, Benito EP, Benoit I, Burger G, Camino LP, Cánovas D, Cerdá-

Olmedo E, Cheng J-F, Domínguez A, Eliáš M, Eslava AP, Glaser F, Gutiérrez G, Heitman J, Henrissat B, Iturriaga EA, Lang BF, Lavín JL, Lee SC, Li W, Lindquist E, López-García S, Luque EM, Marcos AT, Martin J, McCluskey K, Medina HR, Miralles-Durán A, Miyazaki A, Muñoz-Torres E, Oguiza JA, Ohm RA, Olmedo M, Orejas M, Ortiz-Castellanos L, Pisabarro AG, Rodríguez-Romero J, Ruiz-Herrera J, Ruiz-Vázquez R, Sanz C, Schackwitz W, Shahriari M, Shelest E, Silva-Franco F, Soanes D, Syed K, Tagua VG, Talbot NJ, Thon MR, Tice H, de Vries RP, Wiebenga A, Yadav JS, Braun EL, Baker SE, Garre V, Schmutz J, Horwitz BA, Torres-Martínez S, Idnurm A, Herrera-Estrella A, Gabaldón T, Grigoriev IV. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Curr Biol* 26:1577–1584. doi:10.1016/j.cub.2016.04.038

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol Biol Evol* 37:291–294. doi:10.1093/molbev/msz189

Deeds R and. 2014. Crosstalk and the evolution of specificity in two-component signaling. *Proceedings of the National Academy of Sciences* 111:9325–9325. doi:10.1073/pnas.1408294111

Eddy SR. 2011. Accelerated Profile HMM Searches. *Plos Comput Biol* 7:e1002195. doi:10.1371/journal.pcbi.1002195

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. doi:10.1093/nar/gkh340

Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW. 2016. Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. *Mol Biol Evol* 34:247–261. doi:10.1093/molbev/msw223

Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW. 2012. Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors. *Plos Genet* 8:e1003072. doi:10.1371/journal.pgen.1003072

Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. *Nature* 481:394–398. doi:10.1038/nature10724

Galperin MY. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *Bmc Microbiol* 5:35–35. doi:10.1186/1471-2180-5-35

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic biology* 59:307–321. doi:10.1093/sysbio/syq010

Hochberg GKA, Liu Y, Marklund EG, Metzger BPH, Laganowsky A, Thornton JW. 2020. A hydrophobic ratchet entrenches molecular complexes. *Nature* 1–6. doi:10.1038/s41586-020-3021-2

Hochberg GKA, Thornton JW. 2017. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annual review of biophysics* 46:247–269. doi:10.1146/annurev-biophys-070816-033631

Holinski A, Heyn K, Merkl R, Sterner R. 2017. Combining ancestral sequence reconstruction with protein design to identify an interface hotspot in a key metabolic enzyme complex. *Proteins Struct Funct Bioinform* 85:312–321. doi:10.1002/prot.25225

Howard CJ, Hanson-Smith V, Kennedy KJ, Miller CJ, Lou HJ, Johnson AD, Turk BE, Holt LJ. 2014. Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity. *eLife* 3:2104–22. doi:10.7554/elife.04126

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* 1:16048. doi:10.1038/nmicrobiol.2016.48

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. doi:10.1038/s41586-021-03819-2

Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR. 2000. Evolution of Two-Component Signal Transduction. *Mol Biol Evol* 17:1956–1970. doi:10.1093/oxfordjournals.molbev.a026297

Laursen L, Čalyševa J, Gibson TJ, Jemth P. 2020. Divergent Evolution of a Protein–Protein Interaction Revealed through Ancestral Sequence Reconstruction and Resurrection. *Mol Biol Evol* 38:152–167. doi:10.1093/molbev/msaa198

McClune CJ, Alvarez-Buylla A, Voigt CA, Laub MT. 2019. Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature* 574:702–706. doi:10.1038/s41586-019-1639-8

McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW. 2014. Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. *Cell* 159:58–68. doi:10.1016/j.cell.2014.09.003

Mende DR, Letunic I, Huerta-Cepas J, Li SS, Forslund K, Sunagawa S, Bork P. 2017. proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Research* 45:D529–D534. doi:10.1093/nar/gkw989

Mende DR, Letunic I, Maistrenko OM, Schmidt TSB, Milanese A, Paoli L, Hernández-Plaza A, Orakov AN, Forslund SK, Sunagawa S, Zeller G, Huerta-Cepas J, Coelho LP, Bork P. 2019. proGenomes2: an improved database for accurate and consistent habitat, taxonomic and functional annotations of prokaryotic genomes. *Nucleic Acids Research* 210:1518–5. doi:10.1093/nar/gkz1002

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P, Hugenholtz P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 36:996-1004.

Pawson T. 2004. Specificity in Signal Transduction From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems. *Cell* 116:191–203. doi:10.1016/s0092-8674(03)01077-8

Pillai AS, Chandler SA, Liu Y, Signore AV, Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW. 2020. Origin of complexity in haemoglobin evolution. *Nature* 581:480–485. doi:10.1038/s41586-020-2292-y

Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 26:1641–1650. doi:10.1093/molbev/msp077

Siryaporn A, Goulian M. 2008. Cross-talk suppression between the CpxA–CpxR and EnvZ–OmpR two-component systems in E. coli. *Mol Microbiol* 70:494–506. doi:10.1111/j.1365-2958.2008.06426.x

Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. 2008. Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell* 133:1043–1054. doi:10.1016/j.cell.2008.04.040

Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. 2005. Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. *PLoS biology* 3:e334-19. doi:10.1371/journal.pbio.0030334

Starr TN, Picton LK, Thornton JW. 2017. Alternate evolutionary histories in the sequence space of an ancient protein. *Nature* 549:409–413. doi:10.1038/nature23902

Voordeckers K, Brown CA, Vanneste K, Zande E van der, Voet A, Maere S, Verstrepen KJ. 2012. Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *Plos Biol* 10:e1001446. doi:10.1371/journal.pbio.1001446

Wang S, Luo H. 2021. Dating Alphaproteobacteria evolution with eukaryotic fossils. *Nat Commun* 12:3324. doi:10.1038/s41467-021-23645-4

Wheeler LC, Anderson JA, Morrison AJ, Wong CE, Harms MJ. 2018. Conservation of Specificity in Two Low-Specificity Proteins. *Biochemistry-us* 57:684–695. doi:10.1021/acs.biochem.7b01086

Wheeler LC, Harms MJ. 2021. Were Ancestral Proteins Less Specific? *Mol Biol Evol* 38:2227–2239. doi:10.1093/molbev/msab019

Wilson C, Agafonov RV, Hoemberger M, Kutter S, Zorba A, Halpin J, Buosi V, Otten R, Waterman D, Theobald DL, Kern D. 2015. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* 347:882–886. doi:10.1126/science.aaa1823

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24:1586–1591. doi:10.1093/molbev/msm088

Zarrinpar A, Park S-H, Lim WA. 2003. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426:676–680. doi:10.1038/nature02178

855

856

Figure 1

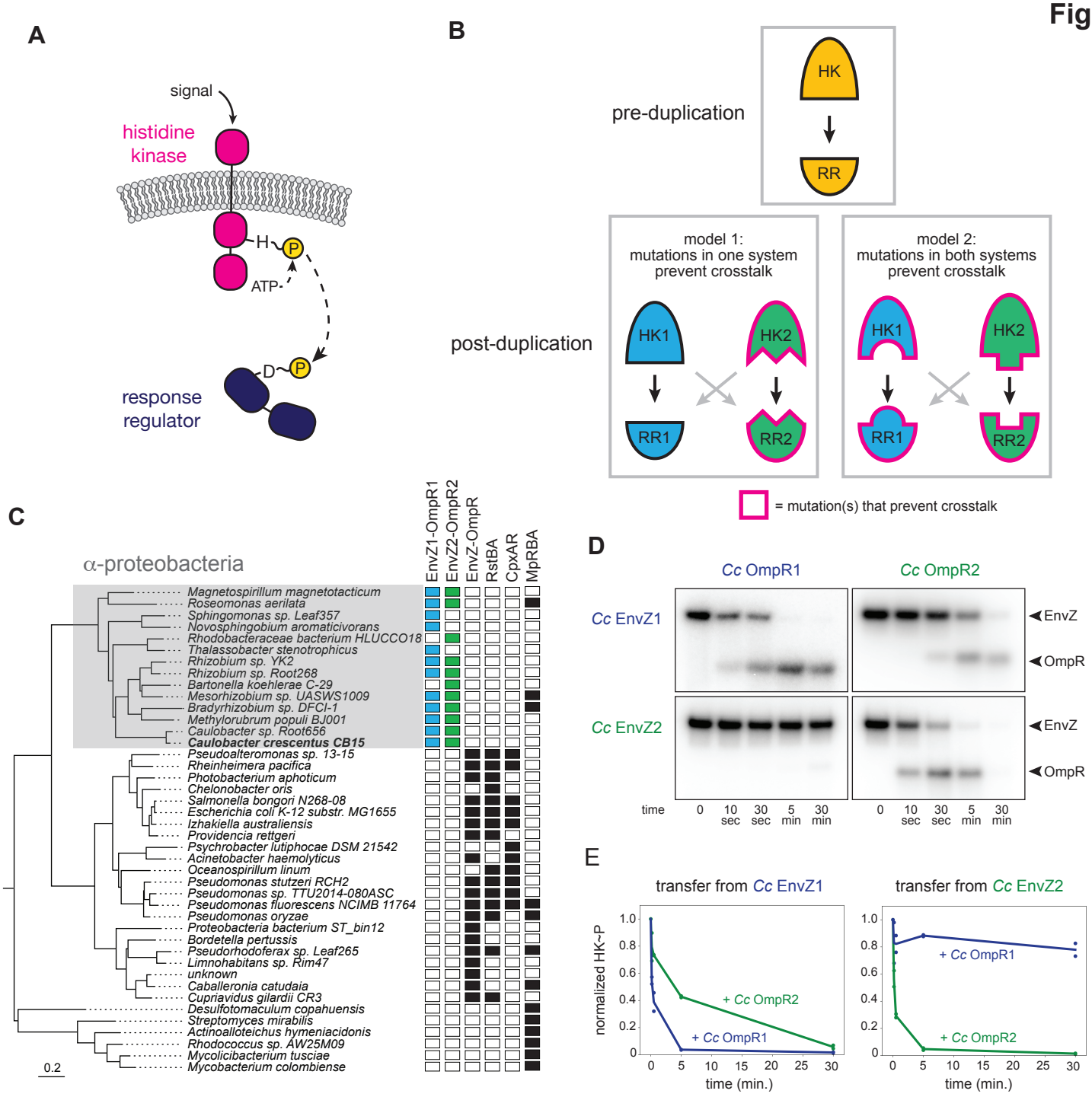

A
signal
histidine kinase
H
P
ATP
D
P
response regulator
B
pre-duplication
HK
RR
post-duplication
model 1: mutations in one system prevent crosstalk
model 2: mutations in both systems prevent crosstalk
HK1
HK2
RR1
RR2
= mutation(s) that prevent crosstalk
C
α-proteobacteria
EnvZ1-OmpR1
EnvZ2-OmpR2
EnvZ-OmpR
RstBA
CpxAR
MprBA
Magnetospirillum magnetotacticum
Roseomonas aerilata
Sphingomonas sp. Leaf357
Novosphingobium aromaticivorans
Rhodobacteraceae bacterium HLUCCO18
Thalassobacter stenotrophicus
Rhizobium sp. YK2
Rhizobium sp. Root268
Bartonella koehlerae C-29
Mesorhizobium sp. UASWS1009
Bradyrhizobium sp. DFCI-1
Methylorubrum populi BJ001
Caulobacter sp. Root656
Caulobacter crescentus CB15
Pseudoalteromonas sp. 13-15
Rheinheimera pacifica
Photobacterium aphoticum
Chelonobacter oris
Salmonella bongori N268-08
Escherichia coli K-12 substr. MG1655
Izhakiella australiensis
Providencia rettgeri
Psychrobacter lutiphocae DSM 21542
Acinetobacter haemolyticus
Oceanospirillum linum
Pseudomonas stutzeri RCH2
Pseudomonas sp. TTU2014-080ASC
Pseudomonas fluorescens NCIMB 11764
Pseudomonas oryzae
Proteobacteria bacterium ST_bin12
Bordetella pertussis
Pseudorhodoferax sp. Leaf265
Limnohabitans sp. Rim47
unknown
Caballeronia catudaia
Cupriavidus gilardii CR3
Desulfotomaculum copahuensis
Streptomyces mirabilis
Actinoalloteichus hymeniacidonis
Rhodococcus sp. AW25M09
Mycolicibacterium tusciae
Mycobacterium colombiense
0.2
D
Cc OmpR1
Cc OmpR2
Cc EnvZ1
Cc EnvZ2
EnvZ
OmpR
time
0
10 sec
30 sec
5 min
30 min
E
transfer from Cc EnvZ1
transfer from Cc EnvZ2
normalized HK~P
+ Cc OmpR2
+ Cc OmpR1
time (min.)

Figure 1 - Figure Supplement 1

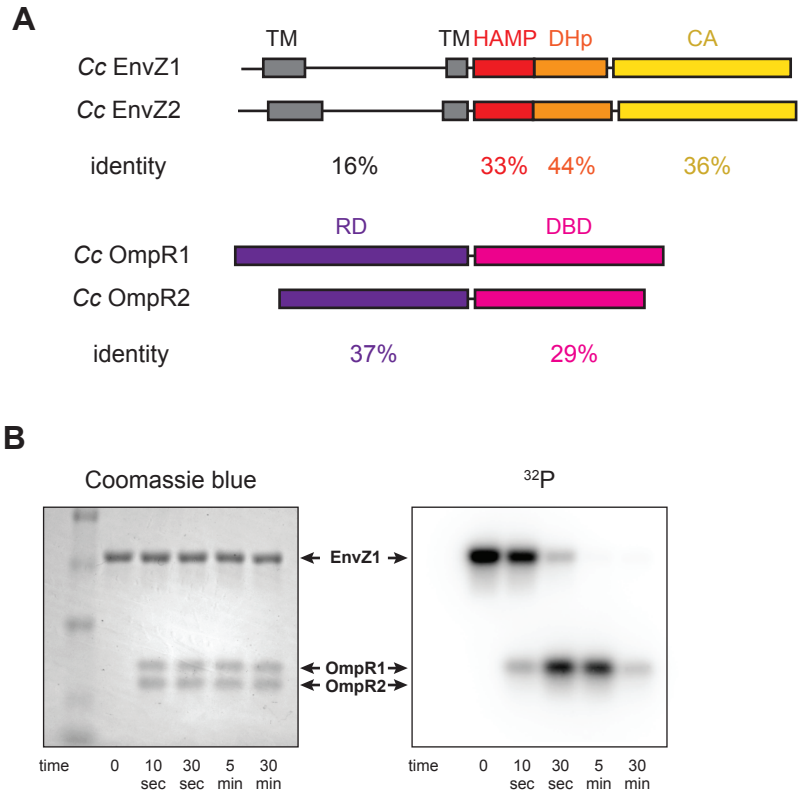


Figure 1 - Figure Supplement 2

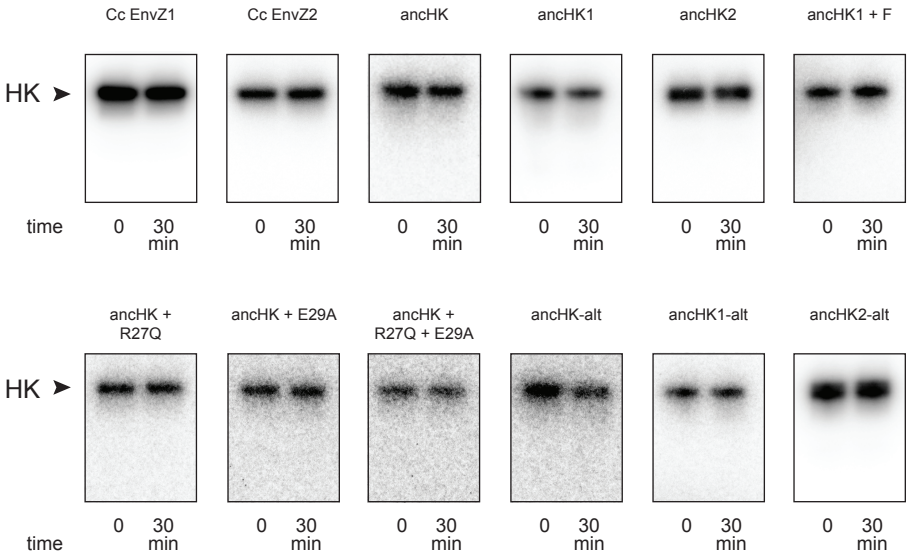


Figure 2

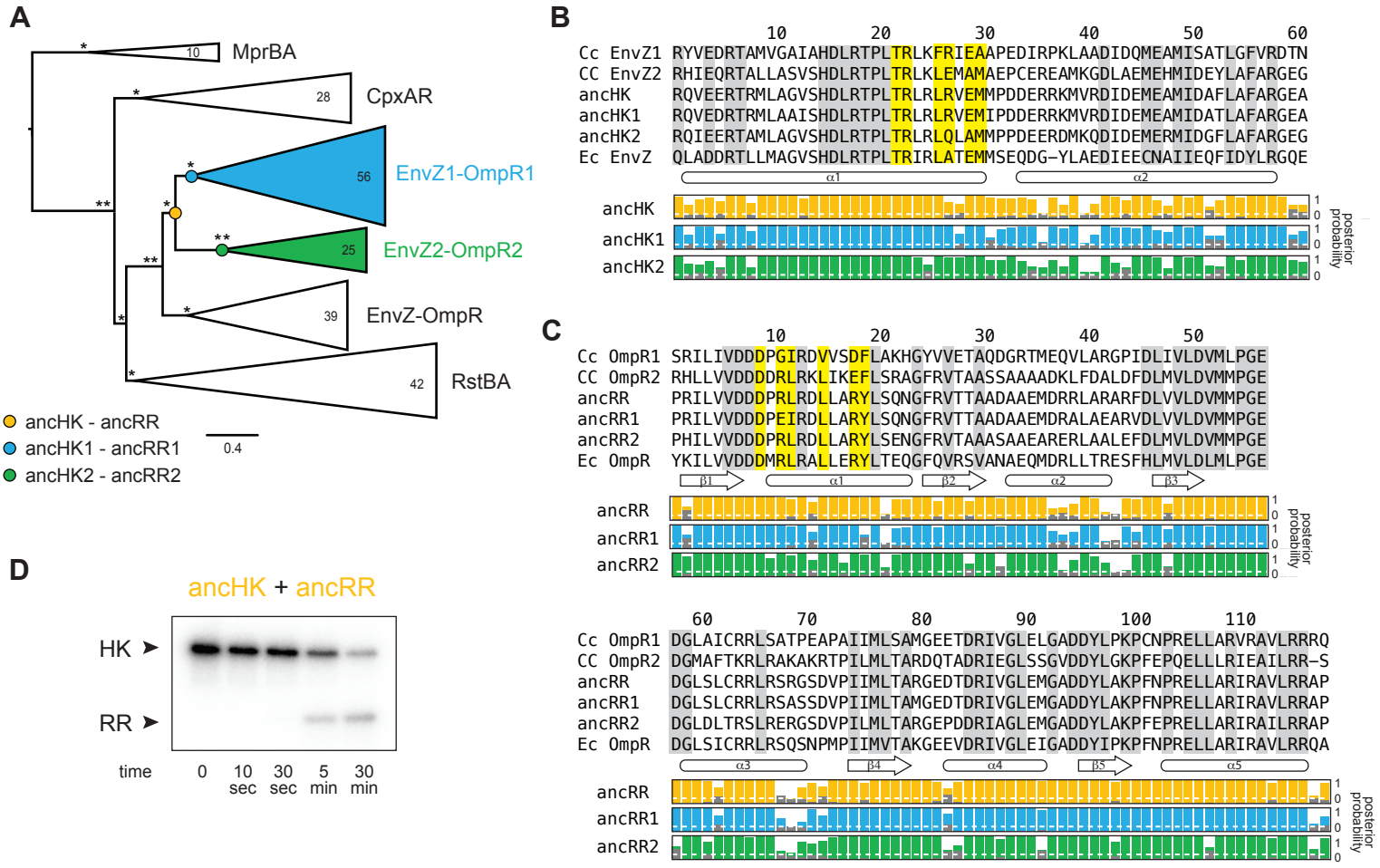

A
MprBA
CpxAR
EnvZ1-OmpR
EnvZ2-OmpR
EnvZ-OmpR
RstBA
ancHK - ancRR
ancHK1 - ancRR1
ancHK2 - ancRR2
0.4
B
Cc EnvZ1 RYVEDRTAMVGAIAHDLRTPLTRLKFRIEAAPEDIRPKLAADIDQMEAMISATLGFVRDTN
CC EnvZ2 RHIEQRTALLASVSHDLRTPLTRLKLEMAMAEPCEREAMKGDLAEMEHMIDEYLAFARGEG
ancHK RQVEERTRMLAGVSHDLRTPLTRLRLRVEMMPDDERRKMVRDIDEMEAMIDAFLAFARGEA
ancHK1 RQVEDRTRMLAAISHDLRTPLTRLRLRVEMIPDDERRKMVRDIDEMEAMIDATLAFARGEA
ancHK2 RQIEERTAMLAGVSHDLRTPLTRLRLQLAMMPPDEERDMKQDIDEMERMIDGFLAFARGEG
Ec EnvZ QLADDRTLLMAGVSHDLRTPLTRIRLATEMMSEQDG-YLAEDIEECNAIIEQFIDYLRGQE
posterior probability
C
Cc OmpR1 SRILIVDDDPGIRDVVSDFLAKHGYVVETAQDGRTMEQVLARGPIDLIVLDVMLPGE
CC OmpR2 RHLLVVDDDDRLRKLIKEFLSRAGFRVTAASSAAAADKLFDALDFDLMVLDVMMPGE
ancRR PRILVVDDDPRLRDLLARYLSQNGFRVTTAADAAEMDRRLARARFDLVVLDVMMPGE
ancRR1 PRILVVDDDPEIRDLLARYLSQNGFRVTTAADAAEMDRALAEARVDLVVLDVMMPGE
ancRR2 PHILVVDDDPRLRDLLARYLSENGFRVTAAASAAEARERLAALEFDLMVLDVMMPGE
Ec OmpR YKILVVDDDMRLRALLERYLTEQGFQVRSVANAEQMDRLLTRESFHLMVLDLMLPGE
Cc OmpR1 DGLAICRRLSATPEAPAIIMLSAMGEETDRIVGLELGADDYLPKPCNPRELLARVRAVLRRRQ
CC OmpR2 DGMAFTKRLRAKAKRTPILMLTARDQTADRIEGLSSGVDDYLGKPFEPQELLLRIEAILRR-S
ancRR DGLSLCRRLRSRGSDVPIIMLTARGEDTDRIVGLEMGADDYLAKPFNPRELLARIRAVLRRAP
ancRR1 DGLSLCRRLRSASSDVPIIMLTAMGEDTDRIVGLEMGADDYLAKPFNPRELLARIRAVLRRAP
ancRR2 DGLDLTRSLRERGSDVPILMLTARGEPDDRIAGLEMGADDYLAKPFEPRELLARIRAILRRAP
Ec OmpR DGLSICRRLRSQSNPMPIIMVTAKGEEVDRIVGLEIGADDYIPKPFNPRELLARIRAVLRRQA
posterior probability
D
ancHK + ancRR
HK
RR
time
0
10 sec
30 sec
5 min
30 min

Figure 2 - Figure Supplement 1

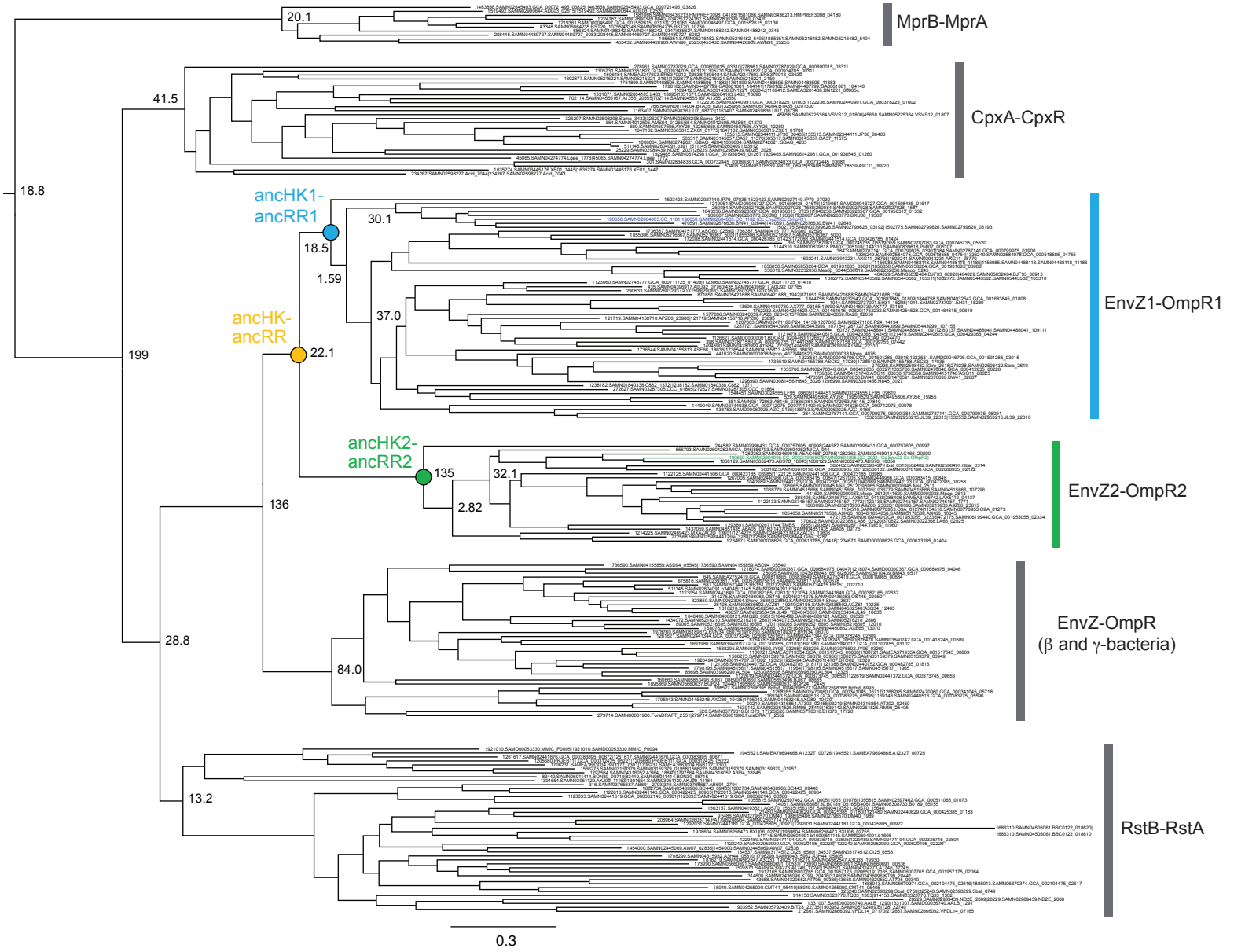


Figure 2 - Figure Supplement 2

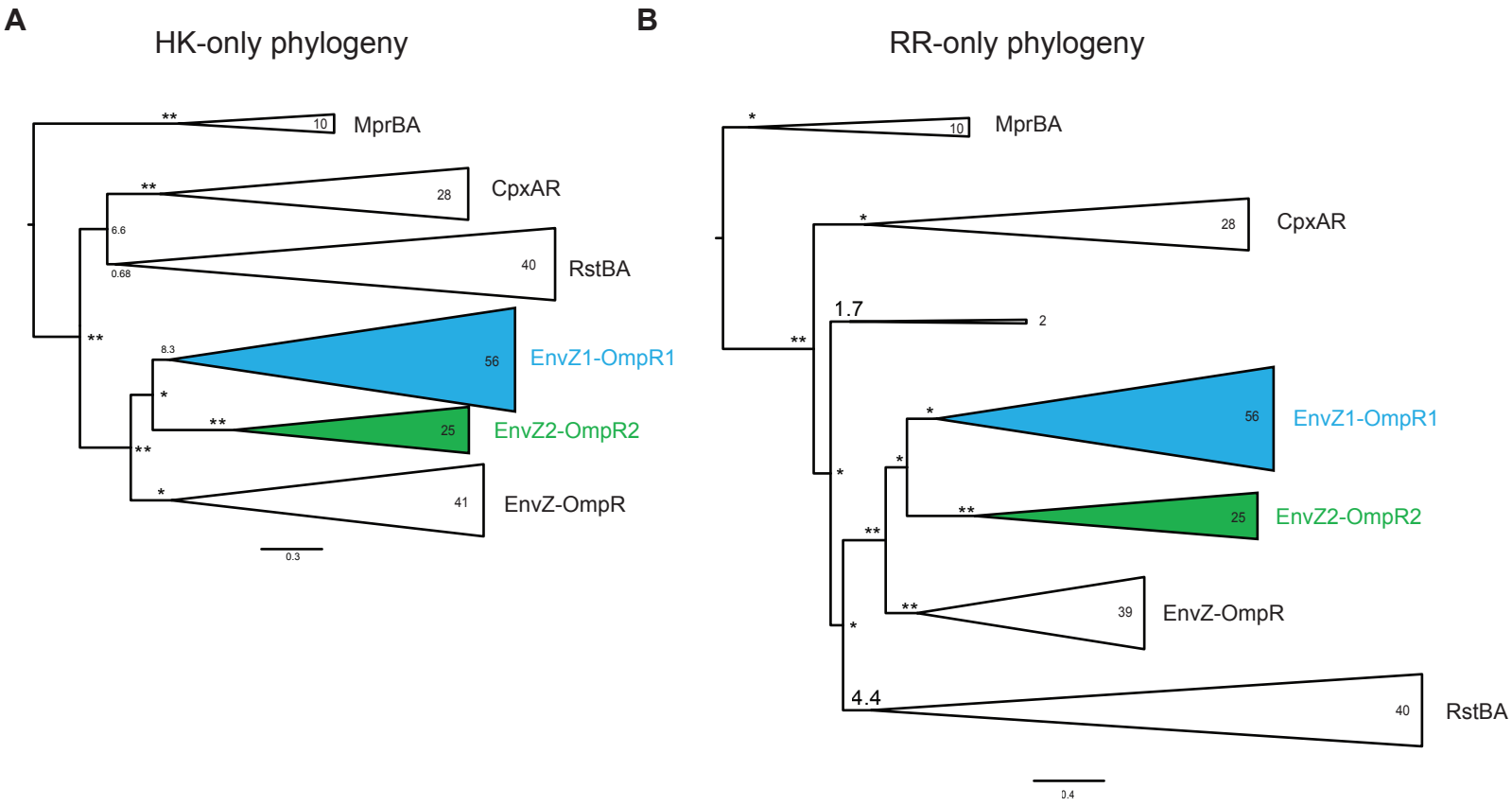


Figure 2 - Figure Supplement 3

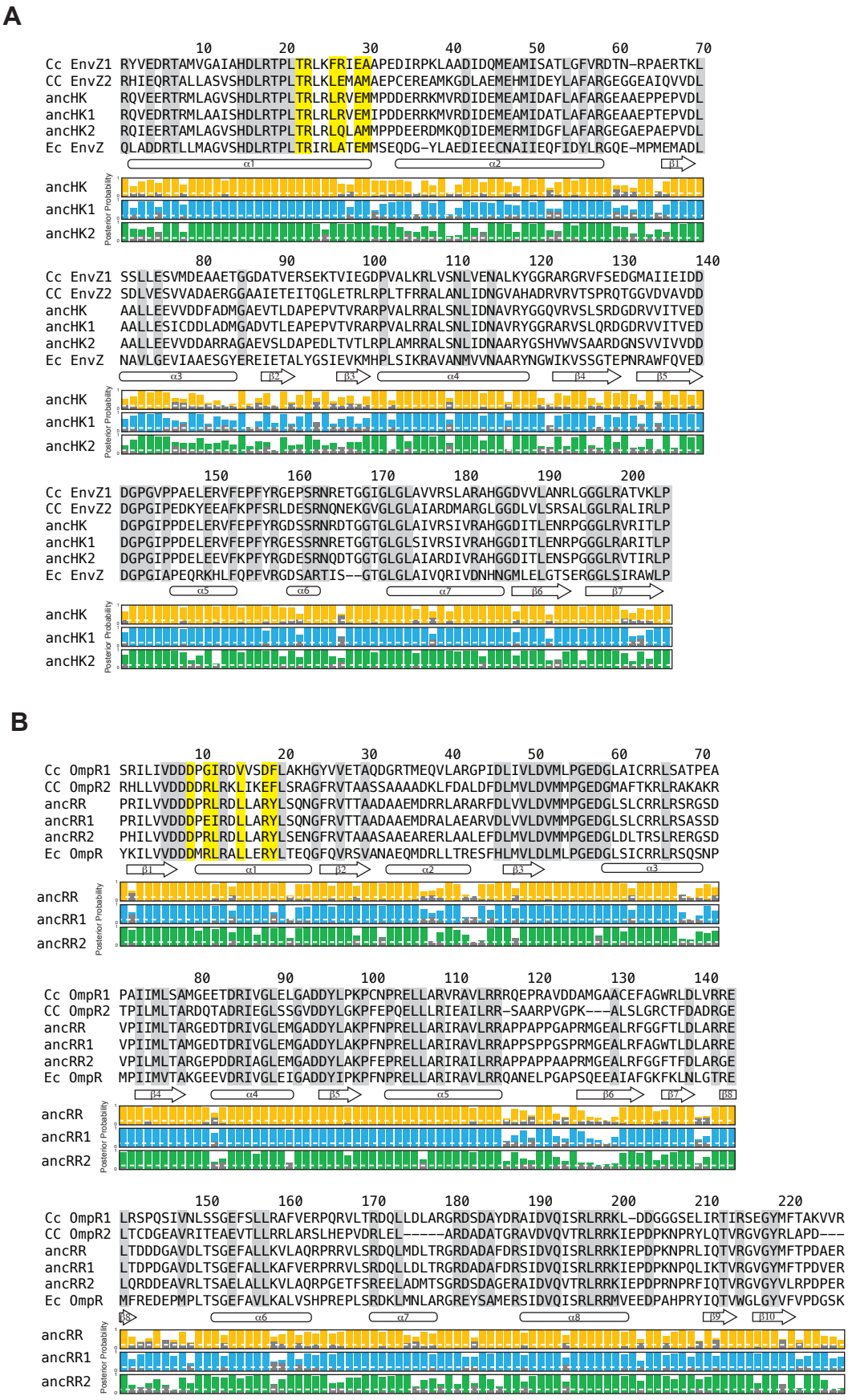
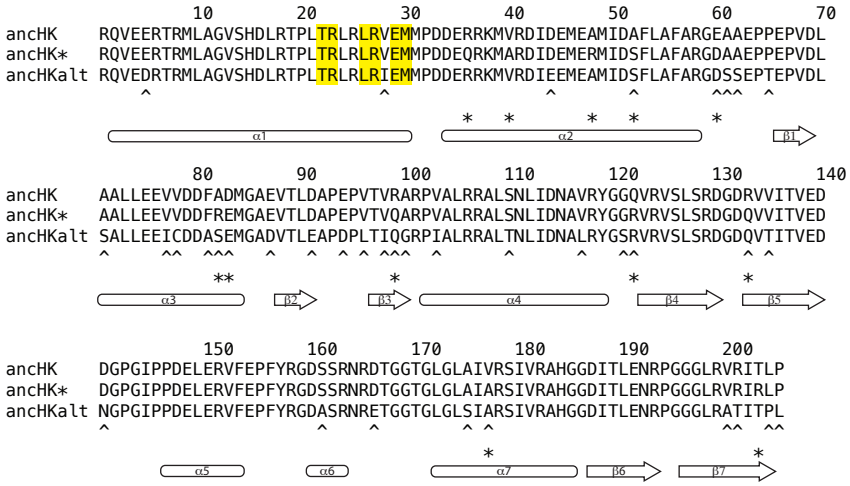
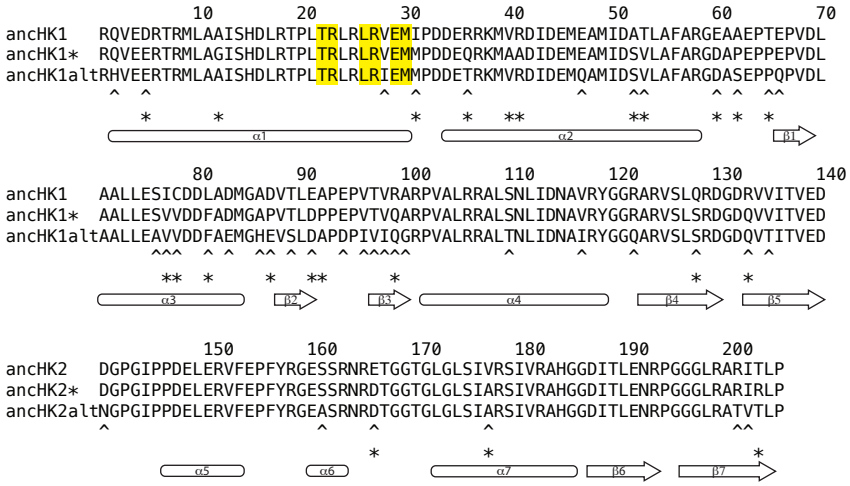


Figure 2 - Figure Supplement 4

A



B



C

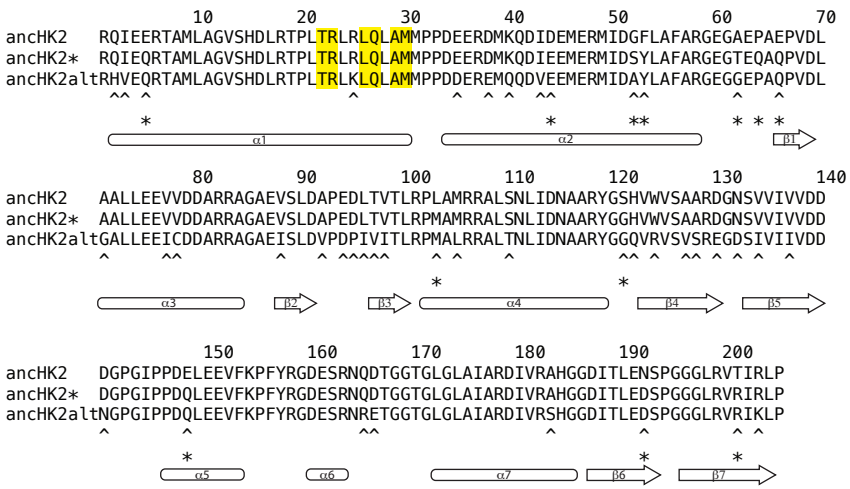


Figure 2 - Figure Supplement 5

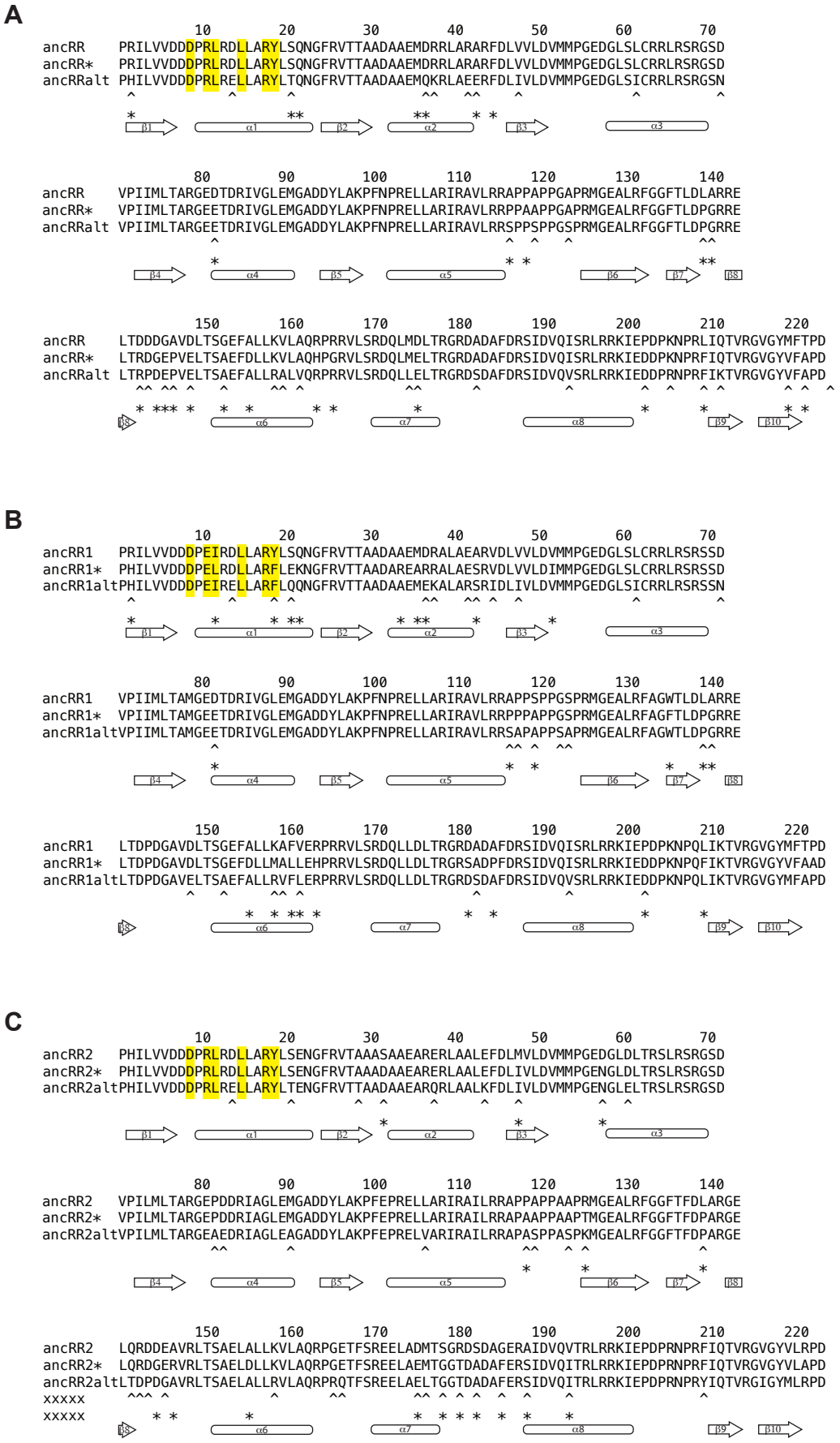


Figure 3

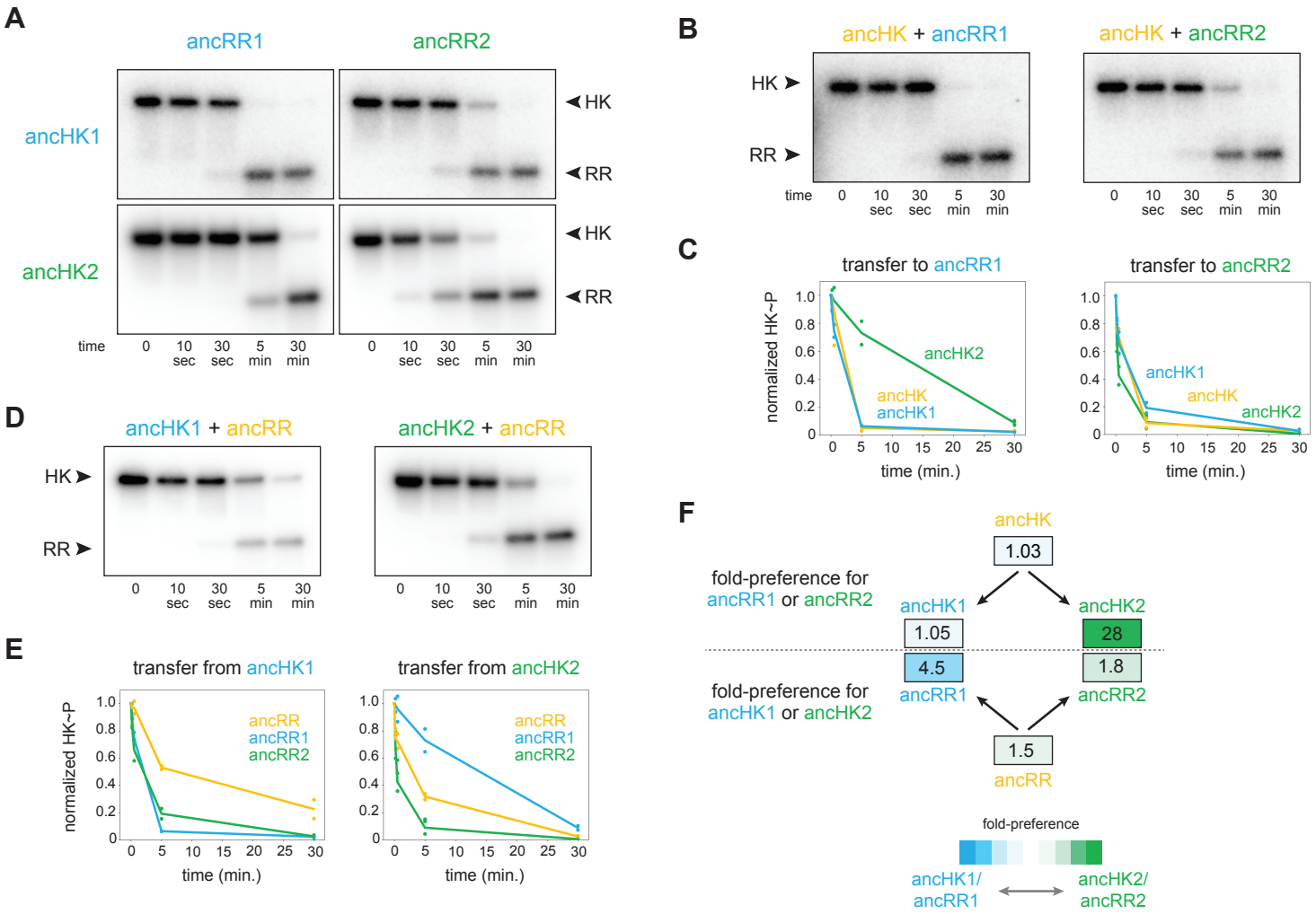
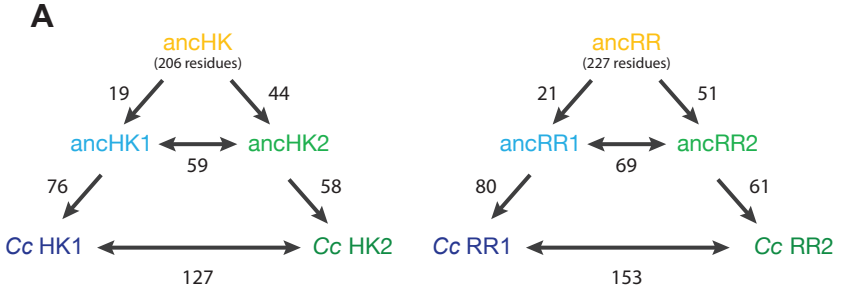

A
ancRR1
ancRR2
ancHK1
ancHK2
HK
RR
time
0
10 sec
30 sec
5 min
30 min
B
ancHK + ancRR1
ancHK + ancRR2
C
transfer to ancRR1
transfer to ancRR2
normalized HK~P
time (min.)
ancHK2
ancHK
ancHK1
D
ancHK1 + ancRR
ancHK2 + ancRR
E
transfer from ancHK1
transfer from ancHK2
ancRR
ancRR1
ancRR2
F
ancHK
1.03
fold-preference for
ancRR1 or ancRR2
ancHK1
ancHK2
1.05
28
4.5
1.8
fold-preference for
ancHK1 or ancHK2
ancRR1
ancRR2
1.5
ancRR
fold-preference
ancHK1/
ancRR1
ancHK2/
ancRR2

Figure 3 - Figure Supplement 2



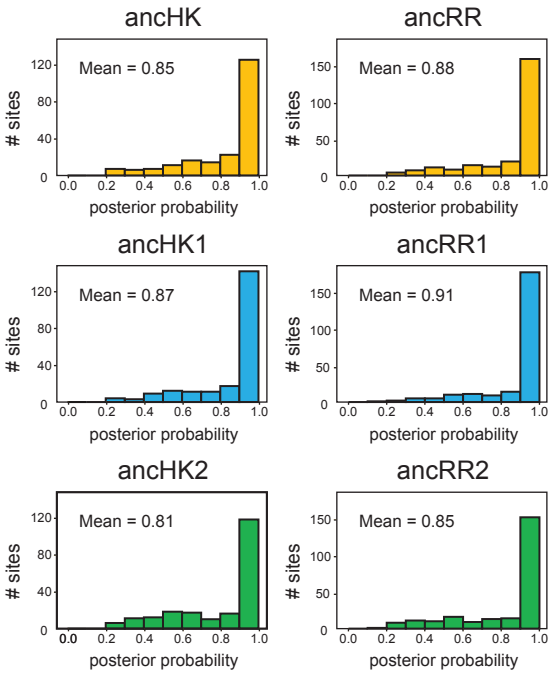
B

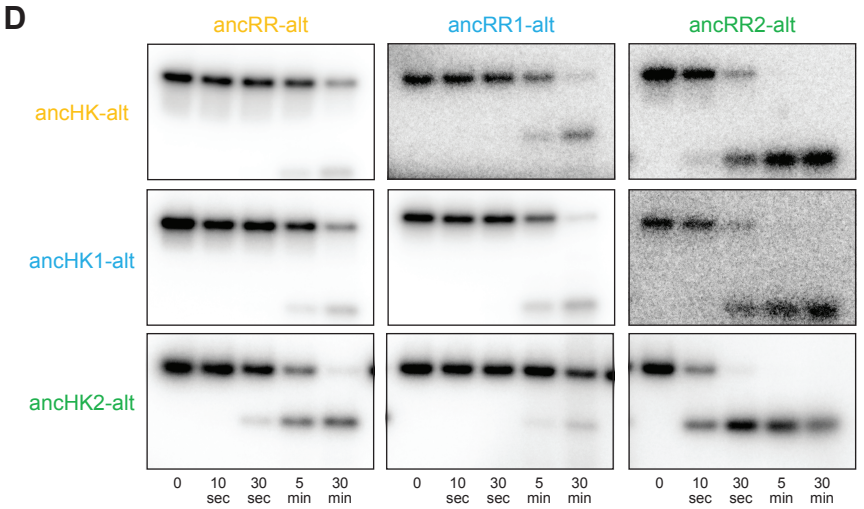
HK Sequence Identity (%)
(DHp + CA domain)

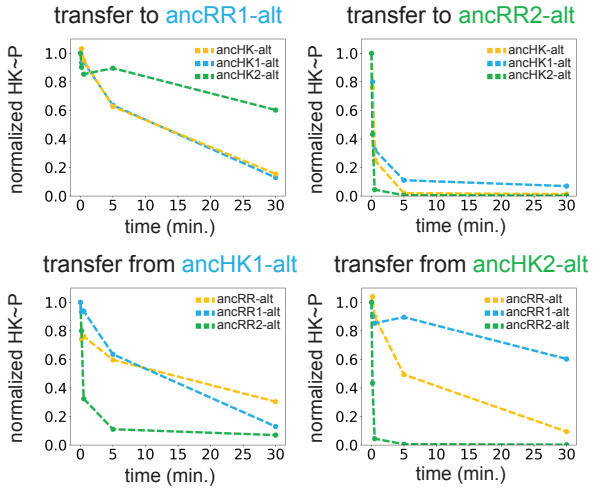
	HK1	HK2	ancHK	ancHK1
HK2	38			
ancHK	54	50		
ancHK1	63	50	91	
ancHK2	47	72	79	71

RR Sequence Identity (%)

	RR1	RR2	ancRR	ancRR1
RR2	32			
ancRR	56	51		
ancRR1	65	42	91	
ancRR2	44	73	78	70







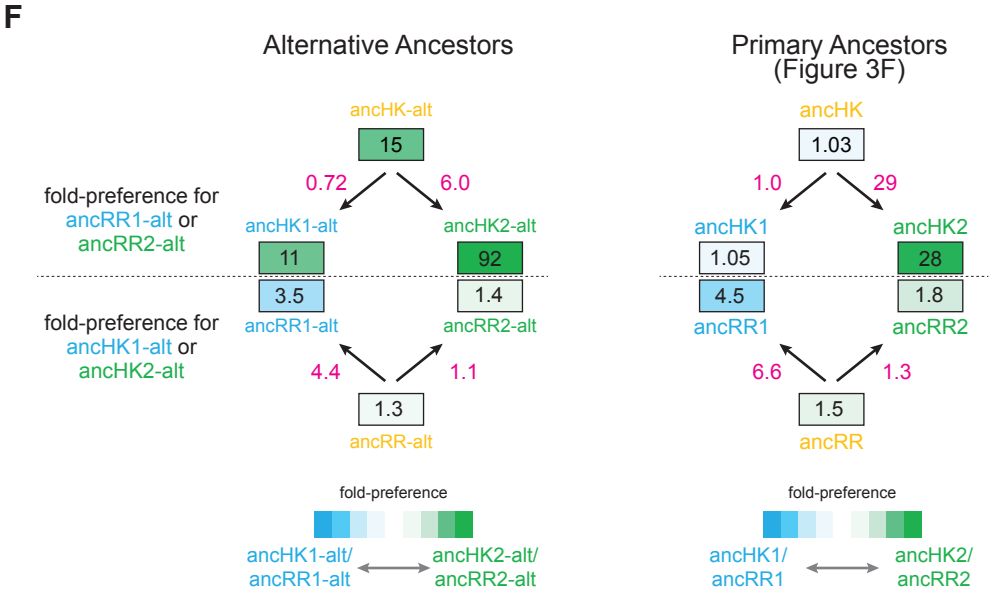
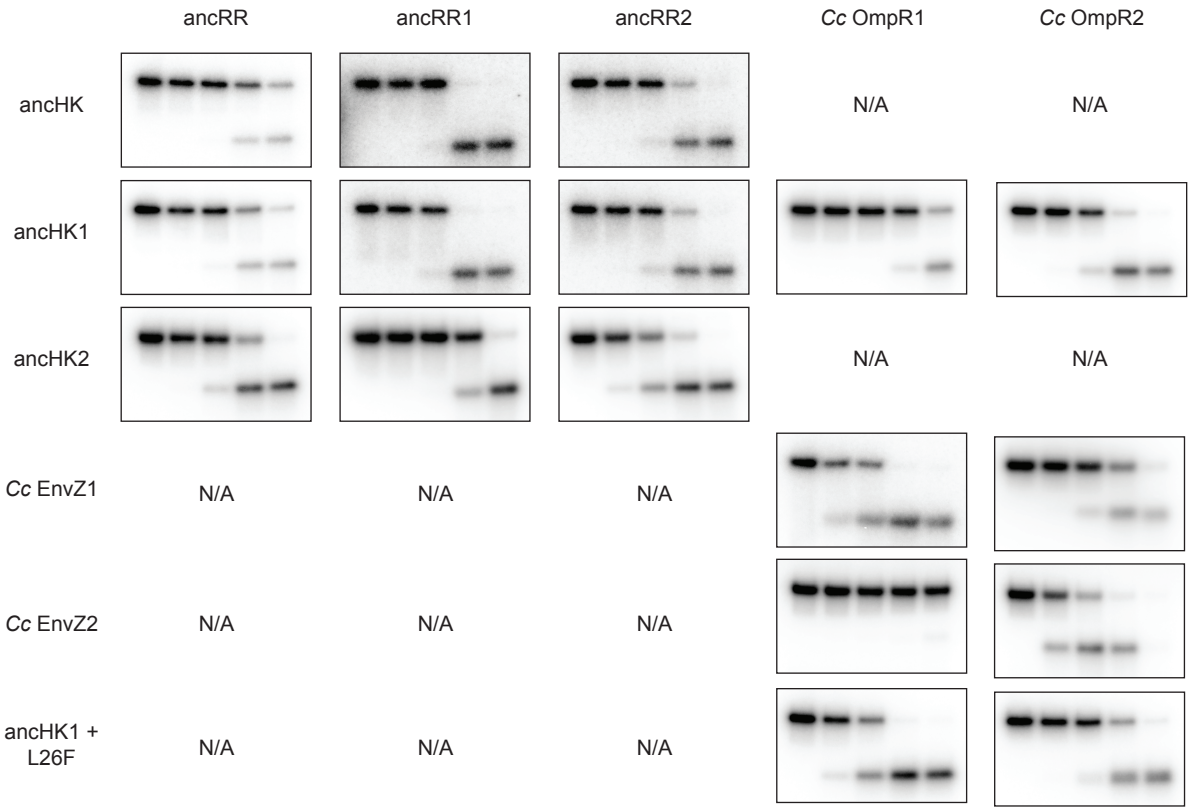


Figure 3 - Figure Supplement 1

A



B

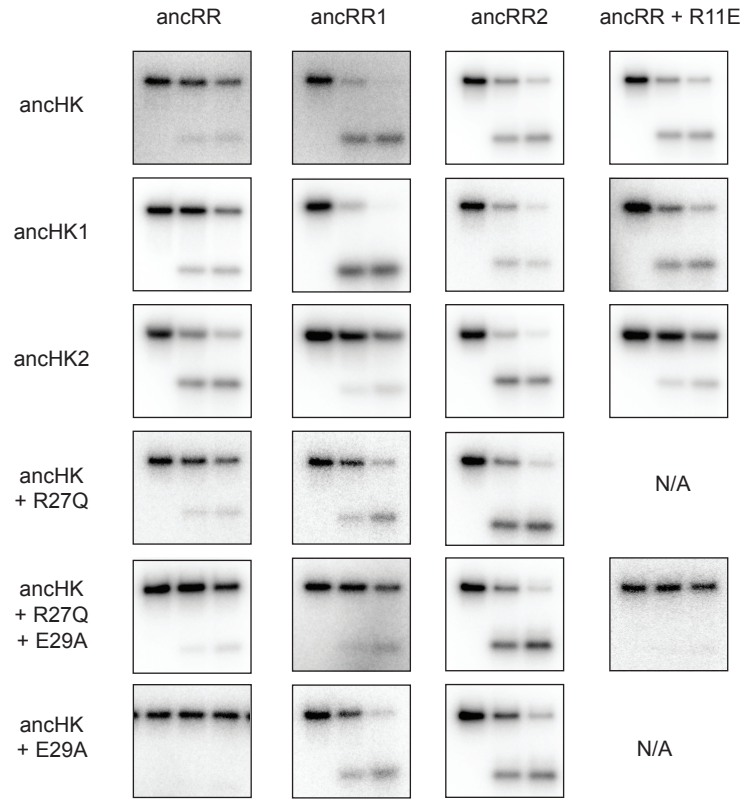


Figure 4

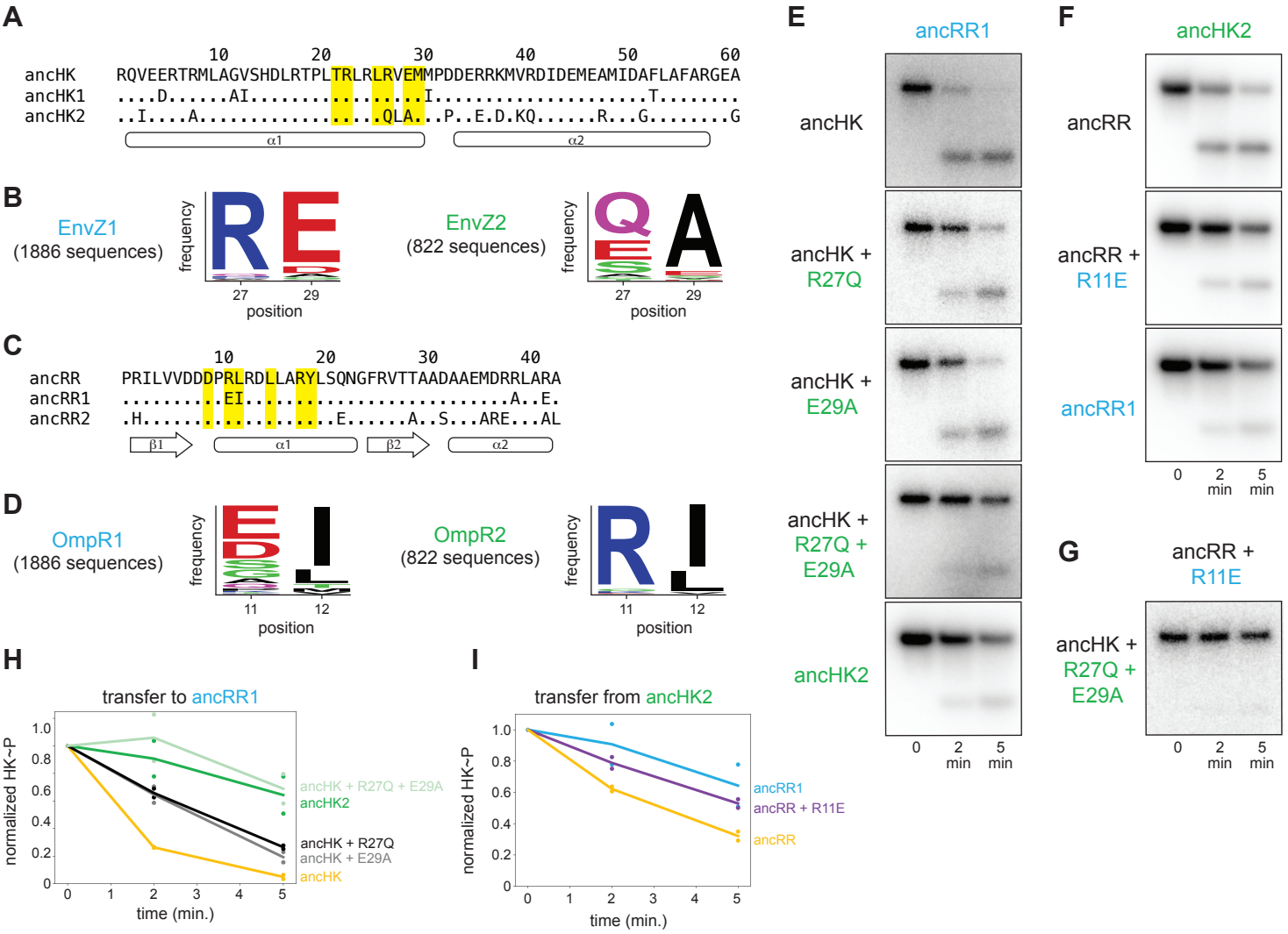

A
ancHK RQVEERTRMLAGVSHDLRTPLTRLRLRVEMMPDDERRKMVRDIDEMEAMIDAFLAFARGEA
ancHK1 ....D......AI.....I.....T.....
ancHK2 ..I....A.....QLA...P..E.D.KQ.....R...G.....G
α1
α2
B
EnvZ1
(1886 sequences)
frequency
position
EnvZ2
(822 sequences)
C
ancRR PRILVVDDDPRLRDLLARYLSQNGFRVTTAADAAEMDRRLARA
ancRR1 .....EI.....A..E.
ancRR2 .H.....E.....A..S...ARE...AL
β1
α1
β2
α2
D
OmpR1
(1886 sequences)
OmpR2
(822 sequences)
E
ancRR1
ancHK
ancHK + R27Q
ancHK + E29A
ancHK + R27Q + E29A
ancHK2
0 2 min 5 min
F
ancHK2
ancRR
ancRR + R11E
ancRR1
G
ancRR + R11E
ancHK + R27Q + E29A
H
transfer to ancRR1
normalized HK~P
time (min.)
ancHK + R27Q + E29A
ancHK2
ancHK + R27Q
ancHK + E29A
ancHK
I
transfer from ancHK2
ancRR1
ancRR + R11E
ancRR

Figure 4 - Figure Supplement 1

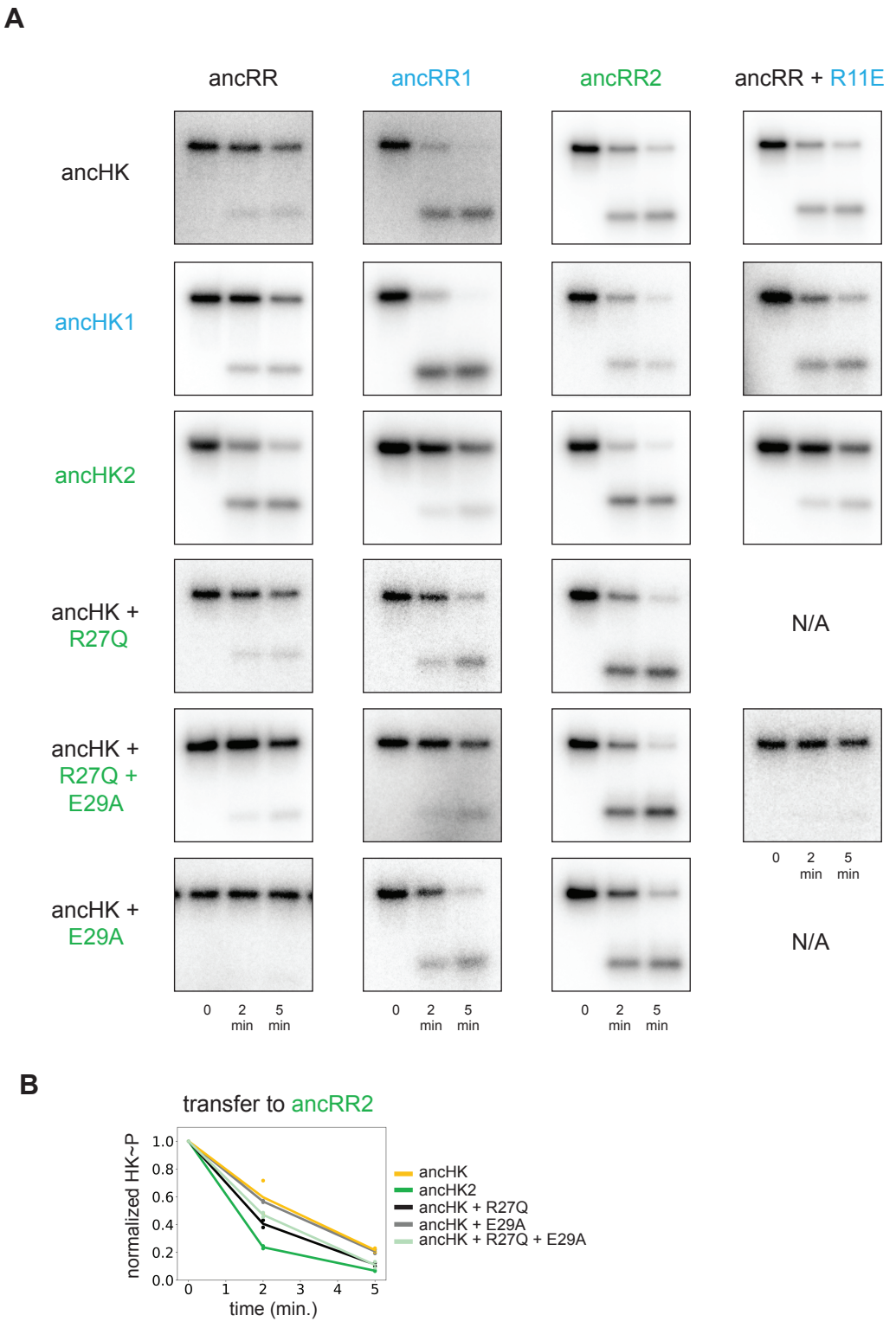


Figure 5

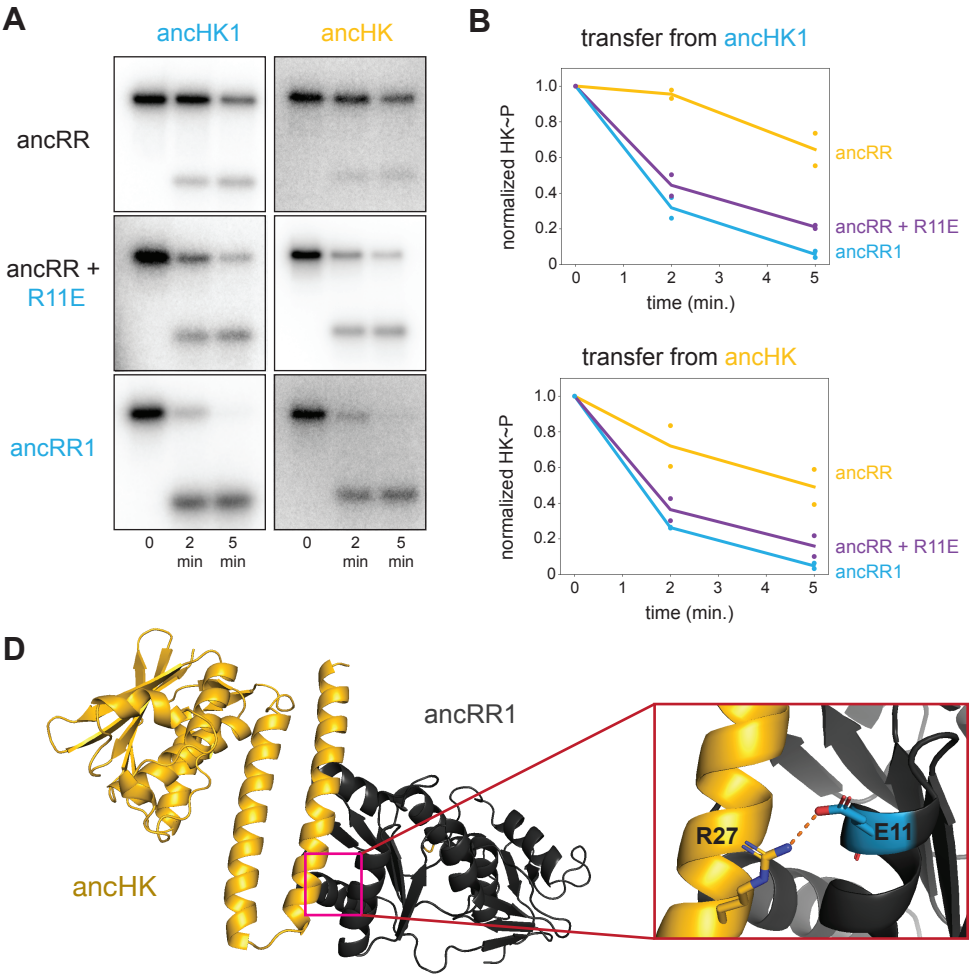

A
ancHK1
ancHK
ancRR
ancRR +
R11E
ancRR1
0
2
min
5
min
B
transfer from ancHK1
normalized HK~P
time (min.)
ancRR
ancRR + R11E
ancRR1
transfer from ancHK
C
HK
pre-duplication
RR
R27Q + E29A
HK1
HK2
post-duplication
RR1
RR2
R11E
R11E
= mutation(s) that prevent crosstalk
= mutation(s) that improve cognate interaction
D
ancRR1
ancHK
R27
E11

Figure 5 - Figure Supplement 1

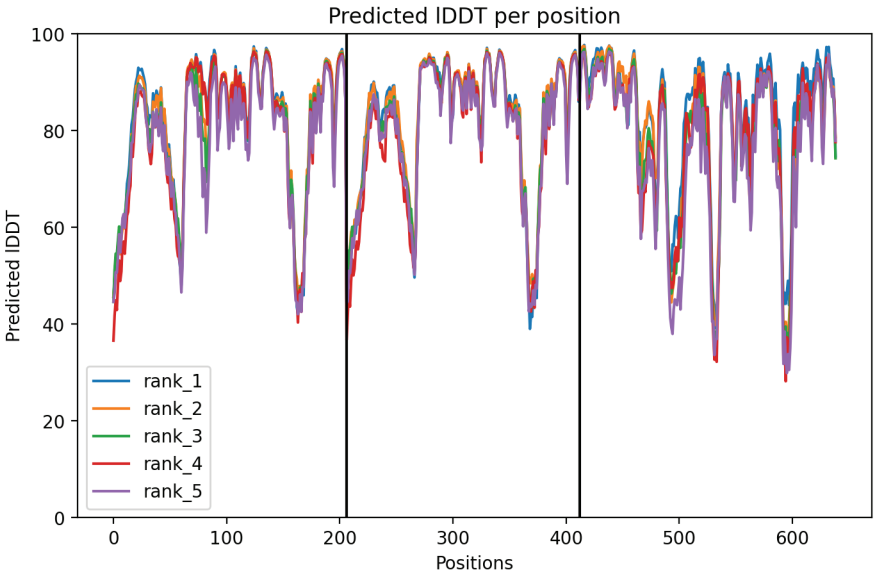
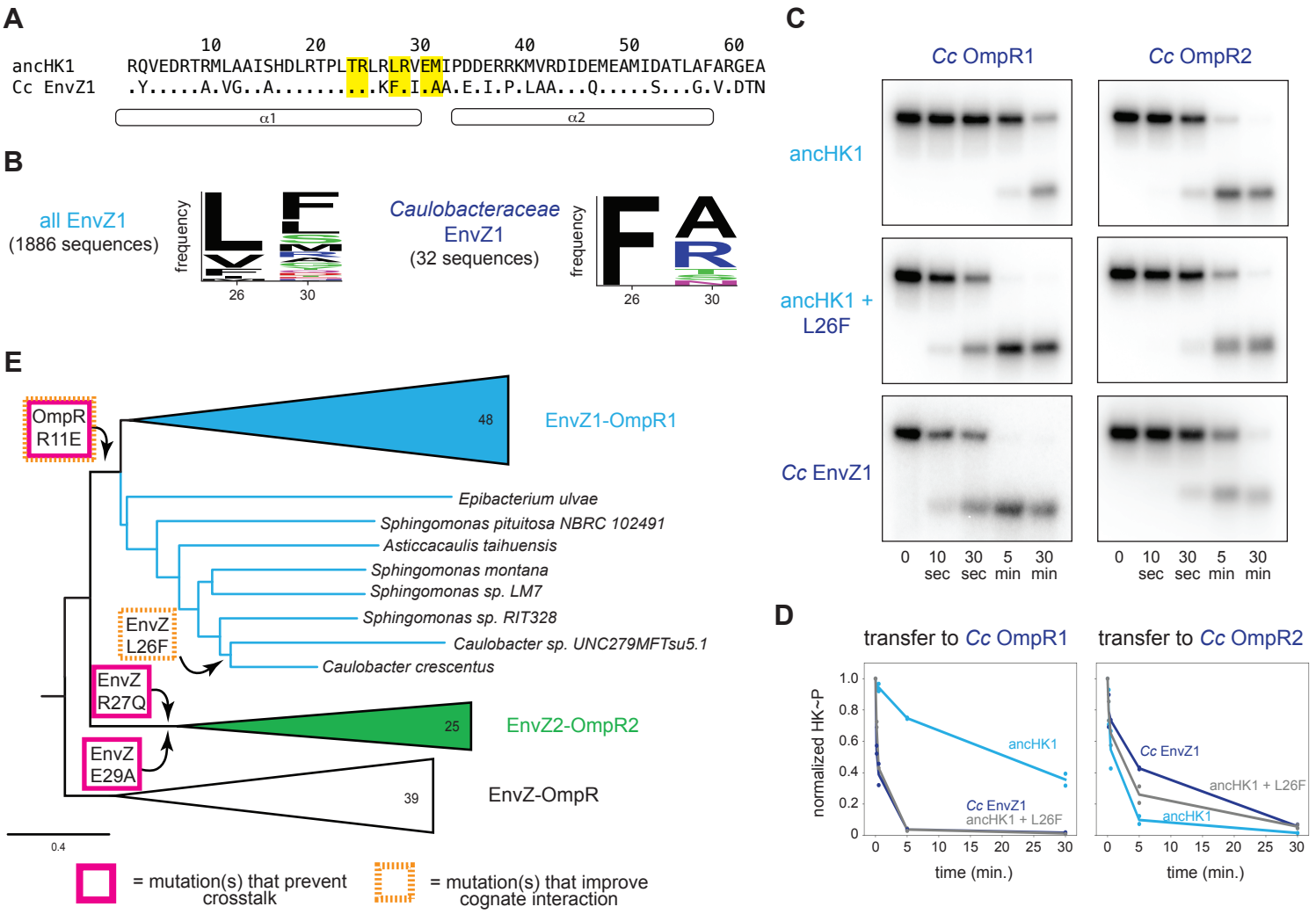


Figure 6


A
10 20 30 40 50 60
ancHK1 RQVEDRTRMLAAISHDLRTPLTRLRLRVEMIPDDERRKMVRDIDEMEAMIDATLAFARGEA
Cc EnvZ1 .Y.....A.VG..A.....KF.I.AA.E.I.P.LAA...Q.....S...G.V.DTN
α1
α2
B
all EnvZ1
(1886 sequences)
frequency
26 30
Caulobacteraceae
EnvZ1
(32 sequences)
frequency
26 30
C
Cc OmpR1
Cc OmpR2
ancHK1
ancHK1 +
L26F
Cc EnvZ1
0 10 sec 30 sec 5 min 30 min
0 10 sec 30 sec 5 min 30 min
D
transfer to Cc OmpR1
transfer to Cc OmpR2
normalized HK~P
ancHK1
Cc EnvZ1
ancHK1 + L26F
time (min.)
Cc EnvZ1
ancHK1 + L26F
ancHK1
time (min.)
E
OmpR
R11E
48
EnvZ1-OmpR1
Epibacterium ulvae
Sphingomonas pituitosa NBRC 102491
Asticcacaulis taihuensis
Sphingomonas montana
Sphingomonas sp. LM7
Sphingomonas sp. RIT328
Caulobacter sp. UNC279MFTsu5.1
Caulobacter crescentus
EnvZ
L26F
EnvZ
R27Q
25
EnvZ2-OmpR2
EnvZ
E29A
39
EnvZ-OmpR
0.4
= mutation(s) that prevent crosstalk
= mutation(s) that improve cognate interaction